

# **Big Datan visualisoinnin kokemus virtuaalitodellisuudessa**

Sakari Hassi

Tampereen yliopisto

Informaatiotieteiden yksikkö

Tietojenkäsittelytieteiden tutkinto-ohjelma

Pro gradu -tutkielma

Ohjaaja: Harri Siirtola

Toukokuu 2018

---

## Tiivistelmä

Tutkielmassa pyrittiin selvittämään vastauksia siihen olisiko virtuaalitodellisuus soveltuva ympäristö Big Datan visualisoimiseen, eli tehostaisiko kokemuksellisempi ympäristö Big Dataksi luokiteltavien datajoukkojen ymmärtämistä. Tutkimuskysymykseen liittyen tutkielmassa haluttiin selvittää, miten käyttäjäkokemus tiedon visualisoinnista eroaa virtuaalitodellisuuden ja työasemaympäristön välillä ja miten käyttäjät kokevat tiedon visualisoinnin kokemuksen virtuaalitodellisuudessa.

Vastausten selvittämiseksi tutkielma aloitettiin Big Datan käsitteen sekä aikaisempien virtuaalitodellisuuteen pohjautuneiden Big Datan visualisointijärjestelmien taustakartoituksella. Aikaisempien visualisointijärjestelmien raportoituja ominaisuuksia peilattiin Big Datan käsitettä vasten ja tehtiin havaintoja siitä, että aikaisemmat ratkaisut ovat huonosti täyttäneet Big Datan käsitteen mukaisia vaatimuksia ja eivät tarjonneet pohjaa tässä tutkielmassa toteutettavia visualisointeja varten. Tutkielman toteutusvaiheessa luotiin kolme visualisointikokonaisuutta, joista luotiin erilliset demot virtuaalitodellisuuteen sekä työasemaympäristöön. Visualisoinnin demot pyrittiin toteuttamaan Big Datan käsitteiden mukaisesti näitä kaikkia kuitenkin saavuttamatta. Tutkielman rajallisilla resursseilla Big Datan asettamista haasteista suurimmaksi koettiin tarpeeksi laajan datamäärän hyödyntämisen sekä Big Datan määritteiden mukaiseen käyttöön soveltuvien tietokantojen löytämisen. Luotuja testijärjestelmiä varten luotiin testisuunnitelma, jonka mukaisesti suoritettiin 10 osallistujan käyttäjätestaus tiedon visualisoinnin kokemusten selvittämiseksi yhtäläisen virtuaalitodellisuustoteutuksen ja työasemaversioiden välillä. Käyttäjätutkimuksissa monet käyttäjät kokivat virtuaalitodellisuuden visualisoinnit kokonaisvaltaisempina kokemuksena ja ympäristö mahdollisti paremman keskittymisen visualisoinnin sisältöön. Osallistujat kuitenkin kokivat, että virtuaalitodellisuuden visualisointien tulisi olla luotuna virtuaalitodellisuuden tarjoamia mahdollisuuksia hyödyntäen, jotta erilaisen ympäristön hyödyntäminen koetaan merkityksellisenä. Tämän lisäksi virtuaalitodellisuudessa käytettyjen vuorovaikutustekniikoiden soveltuvuus sekä järjestelmän käytön sujuvuus korostuivat huomiota herättäneinä tekijöinä.

**Avainsanat ja -sanonnat:** Big Data, virtuaalitodellisuus, visualisointi, käyttäjäkokemus.

# SISÄLLYSLUETTELO

1. JOHDANTO .....	1
2. BIG DATA .....	2
2.1 Big Datan määrittely.....	2
2.2 Big Datan kuudes V – Visualisointi .....	4
2.3 Big datan hyödyt ja ongelmat .....	6
3. TIEDON VISUALISOIMINEN .....	9
3.1 Visualisointi yleisesti.....	9
3.2 Visualisoinnin tyypit .....	10
3.2.1 Parallel coordinates .....	12
3.2.2 Star coordinates.....	13
3.2.3 Tree map.....	15
3.3 Visualisoinnin työkalut.....	16
3.4 Visualisoinnin prosessi .....	17
3.5 MapReduce .....	19
4. VIRTUAALITODELLISUUDEN HYÖDYNTÄMINEN .....	21
4.1 Virtuaalitodellisuus aikaisemmin .....	21
4.2 Big Datan visualisointijärjestelmät.....	22
4.3 Yhteenveto aikaisempien järjestelmien pohjalta .....	25
4.4 HTC Vive .....	26
5. TESTIJÄRJESTELMÄN KEHITYS .....	28
5.1 Twitter-viestien visualisoiminen maailmankartalla.....	29
5.1.1 Toiminnollisuus .....	30
5.2 Twitter-viestitietojen visualisoiminen pylväsdiagrammeilla ja graafilla .....	31
5.2.1 Toiminnollisuus .....	32
5.3 Big Datan hyödyntäminen ja sen ongelmat testijärjestelmissä .....	33
6. TESTIJÄRJESTELMIEN KÄYTTÄJÄTESTAUS .....	37
6.1 Testisuunnitelma.....	37
6.2 Tulokset.....	41

6.2.1 Osioiden käyttökokemuksen arviot.....	42
6.2.2 Loppukysymysten vastaukset .....	45
7. YHTEENVETO.....	48
LÄHDELUETTELO .....	51
LIITTEET .....	58
Taustatietolomake.....	58
Käyttäjätutkimuksen tehtävät:.....	59
Käyttäjätutkimuksen haastattelukysymyslomakkeet .....	60
Käyttäjätutkimuksen tulokset ja taulukot .....	62
Osallistujat .....	62
Vastaukset käyttäjäkohtaisesti .....	62

# 1. JOHDANTO

Tieto on aina ollut merkittävässä roolissa ja sillä on aina nähty olevan strategista arvoa. Nykypäivän tärkein varanto ei ole enää materia, vaan varallisuus perustuu aineettomaan omaisuuteen, jota ohjataan tiedolla. Myös Buyya ja kumppanit [2009] toteavat tutkimuksessaan, että tietojenkäsittelyn kautta tuotettu tieto tulee olemaan viides hyödyke sähkön, veden, bensiinin ja teleliikenteen lisäksi. Tiedon merkittävyyden kasvuun liittyen Big Data on ollut jo muutamia vuosia kuuma puheenaihe tutkimus- ja yritysmaailmassa. Alati kasvava luodun tiedon määrä ja sensoriverkkojen kehitys on johtanut tilanteeseen, jossa yhä useammasta fyysisen maailman tapahtumasta jää digitaalinen jälki. Tätä tietoa on nyt yhä laajemmin alettu arvostamaan ja käyttämään osana päätöstentekoprosesseja. Big Datan ongelmana on kuitenkin sen käyttöön valjastaminen. Tiedon tarkasta sisällöstä ei voida olla varmoja ja tiedon sisäisten yhteyksien löytäminen sekä analysoiminen ovat työläitä prosesseja. Näiden tekijöiden lisäksi tiedon käsittely vaatii paljon säilytyskapasiteettia sekä Big Datan käsittelyyn luodun ohjelmistoarkkitehtuurin. Big Datan käsittelyn ongelmat huomioiden, Brenton Faber totesi vuoden 2013 Association of Teachers of Technical Writing -konferenssissa meidän teknologisten mahdollisuuksien kerätä tietoa ylittävän analyyttisen kapasiteetin luoda uutta ja merkityksellistä tietoa tämän aineiston pohjalta [Faber, 2013].

Kasvavien datamäärien käsittelyyn ja hyödyntämiseen on pyritty etsimään ratkaisua uusista teknologia-alan innovaatioista. Virtuaalitodellisuus ei ole uusi teknologia, vaikka virtuaalitodellisuuteen pohjautuvat ratkaisut ovat yleistyneet vasta muutamia vuosia sitten. Virtuaalitodellisuutta on alettu hyödyntää esimerkiksi peleissä sekä markkinointityökaluna. Virtuaalitodellisuutta on myös esitetty ratkaisuna alati kasvavien ja kompleksisten datakokonaisuuksien esittämiseen. Dataa virtuaalitodellisuudessa visualisoivia järjestelmiä ei kuitenkaan ole toistaiseksi juuri saatavilla. Tässä tutkimuksessa tarkoituksena onkin arvioida virtuaalitodellisuuden soveltuvuutta datan visualisointiin ja mitä haasteita tai mahdollisuuksia virtuaalitodellisuus tuo verrattuna normaaliin työpöytäympäristöön. Toisena tutkimuskysymyksenä on tarkoitus selvittää miten käyttäjien käyttökokemukset eroavat virtuaalitodellisuudessa ja työasemaympäristössä toteutettujen visualisointien välillä. Yhteenvetona pyritään avaamaan syitä sille miksi ennakkonäkemyksistä huolimatta virtuaalitodellisuuden ratkaisut eivät ainakaan toistaiseksi ole saavuttaneet suurta suosiota Big Datan visualisoinnissa.

*"Our technical ability to gather data exceeds  
our analytical capacity to make meaning from this data."  
Brenton Faber, 2013*

## 2. BIG DATA

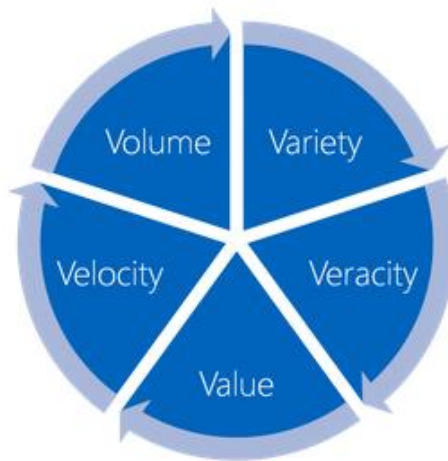
### 2.1 Big Datan määrittely

Big Data on käsitteenä yhä varsin uusi ja sen määrittelyssä ei ole vielä saavutettu täyttä konsensusta. Joidenkin mielestä käsite kuvaa vain suurta datajoukkoa, kun taas joidenkin mielestä käsite on huomattavasti moniulotteisempi. Määritysten kehittäminen on erityisesti keskittynyt kuvaamaan kokonaisuutta, joka olisi tarpeeksi suuri, jotta se laskettaisiin Big Dataksi. Vuonna 2014 Berkeleyn School of Information [Dutcher, 2014] teki kyselytutkimuksen ja pyrki selvittämään eri alojen asiantuntijoiden vastausten pohjalta yhteisen määrittelyn Big Datalle, mutta sai vastaukseksi 40 jokseenkin toisistaan poikkeavaa määritelmää. Yhteneväisyytenä näille määrittelyille pystyttiin kuitenkin toteamaan datan massiivisuus, monimuotoisuus ja tiedon muutosnopeus. Nykyään Big Datan määrittelyssä poikkeuksetta esiintyykin ainakin kolmen V:n määritelmä: tiedon määrä (*Volume*), tiedon nopeus (*Velocity*) ja tiedon monimuotoisuus (*Variety*). Näitä voidaankin pitää pohjamäärittelynä Big Datan kuvaamiseen. Laney [2001] alusti Big Datan kolmen V:n määritelmän kuvatessaan yrityksen tiedonhallinnan kasvavia ongelmia toistaiseksi vielä kuitenkin mainitsematta Big Datan käsitettä osana esitystään. Big Data -käsitettä käytettiin ensimmäisen kerran Coxin ja Ellsworthin [1997] tekemässä tutkimuksessa, joka tämän tutkielman tavoin liittyi isojen tietomäärien visualisoimiseen. Nykyisen pohjamäärittelyn mukaisen Big Data -käsitteen kuitenkin esitteli vasta vuonna 2005 O'Reilly-median jäsen Roger Magoulas, pyrkiessään kuvaamaan suurta tiedon määrää. Voidaan siis todeta, että Big Data määrittelemisessä ei ole toistaiseksi saavutettu täyttä konsensusta. Seuraavat kolme tekijää on kuitenkin vahvasti kartoitettu aihetta käsittelevissä tutkimuksissa ja ne voidaan nähdä merkittävimpinä tekijöinä Big Datan käsittelyssä.

- *Volume*, Tiedon määrä on kasvanut niin suureksi, että käsiteltävä data ei enää mahdu analysoinnissa käytettävien tietokoneiden muistiin. Tästä syystä Big Datan käsittely vaatii erityisesti sille luodun järjestelmäympäristön. Esimerkkinä International Data Corporation [2014] toteaa ennustuksessaan datan määrän tuplaantuvan kahden vuoden välein vuonna 2015 datan määrä ollessa 8 zettatavua ( $10^{21}$ ). Tämän kaavan pohjalta vuonna 2020 dataa on ennustettu olevan jo 44 zettatavua.
- *Velocity*, kuvaa tiedon muuttumisen ja liikkumisen nopeutta. Tieto on nykyään reaaliaikaista ja tiedon analysointi tulisi suorittaa nopeasti sen tuottamisen jälkeen, sillä muuten tieto menettää nopeasti merkityksensä. IBM:n teettämien arvioiden mukaan vuonna 2018

globaalisti internet-protokollaa hyödyntävien laitteiden kaista on 50 000 gigatavua sekunnissa [IBM, 2013].

- *Variety*, Tiedon sisältö on muuttunut pelkistä numeroista ja tekstistä rakenteettomaan suuntaan eri laitteiden tuottaessa rakenteellisesti erilaista dataa. Tällöin käsiteltävä tieto vaatii osakseen metatietoja, joiden avulla tiedon sisältö pystytään kuvaamaan. Tiedon sisällön tarkka ymmärtäminen onkin yksi avaintekijöistä Big Dataa hyödynnettäessä.



Kuva 1. Big datan viisi V:tä

(<https://www.omnivex.com/company/blog/what-is-big-data>)

Laney'n määrittelemien alkuperäisen kolmen V:n lisäksi on useissa tutkimuksissa [Hadi *et al.*, 2014; Rajaraman, 2016] esitelty yhä kaksi jatkomäärittelevää V:tä: *Veracity* ja *Value* (Kuva 1).

- *Veracity*, Kerätty data ei välttämättä ole todenmukaista. Data saattaa olla huonosti organisoitua tai puutteellista, jolloin sen sisältöön ei voida täysin luottaa ja tietoa ei voida käyttää osana luotettavaa päätöksentekoa. Erityisesti autonomisten, tietokonealgoritmien avulla ohjattujen, *Cyber-Physical-Systems (CPS)* järjestelmien kohdalla sensoridatan epäluotettavuus on arvioitu suurimmaksi verkon toimintaa estäväksi tekijäksi [Tolle *et al.*, 2005; Szewczyk *et al.*, 2004].
- *Value*, Tieto itsessään on arvotonta, ellei sitä pystytä prosessoinnin kautta hyödyntämään päätöksenteossa ja toiminnassa. Big Datan kohdalla ongelma on tiedon suuri määrä, mikä tulisi prosessoida ennen kuin siitä voidaan luoda arvoa. Perinteinen työasema ei itsessään riitä prosessoinnin suorittamiseen, vaan käsittely vaatii hajautetun tiedonhallinta – ja prosessointiympäristön. Tällä hetkellä parhaimman alustan tähän tarjoaa *Hadoop* (*Hadoop*

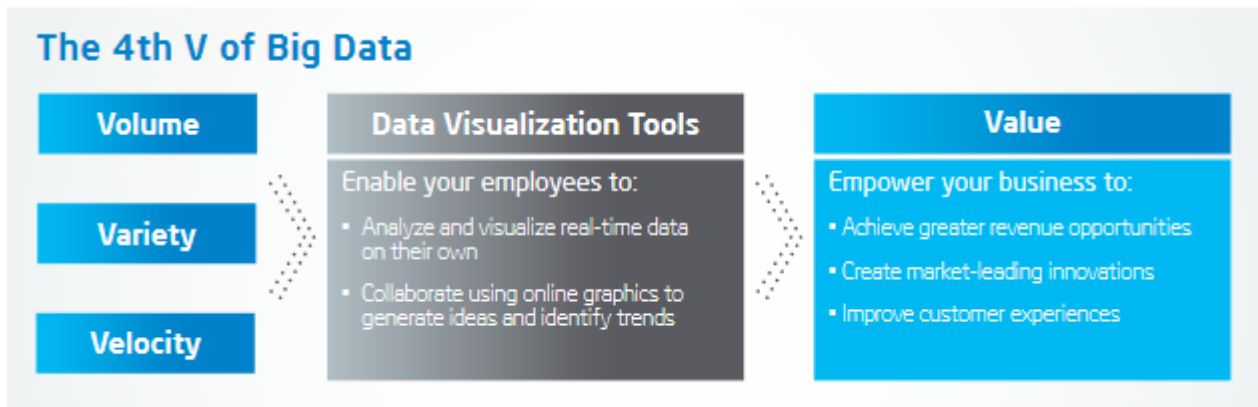
*Distributed File System*) sekä *MapReduce* -ohjelmointimalli (*Distributed computation framework*) [Shvachko *et al.*, 2010].

Tässä tutkimuksessa Big Datan määritelmä nähdään kuitenkin Moninon [2016] antaman yleisemmän määrittelyn mukaisesti: Big Data on pragmaattinen datan analysoinnin muutos, jossa käytetään hyvin tunnettuja kaavoja, joiden avulla tietoon piilotetut relaatiot ja yhteydet saadaan esille. Näitä löydettyjä yhteyksiä voidaan hyödyntää korkeamman tiedon johtamisessa ja johdetun arvon käyttämisessä osana päätöksentekoprosessia. Tämä voidaan nähdä radikaalina muutoksena yritysten toimintatavoissa, joissa yritys pyrkii tuottamaan lisäarvoa itse keräämänsä ja tuottamansa tiedon pohjalta. Yritysten suurin ongelma ei ole enää päättää pitäisikö sen esimerkiksi julkaista uusi tuote markkinoille vaan hyödyntää jo olemassa olevaa tietoa ratkaisujen tekemiseksi.

## **2.2 Big Datan kuudes V – Visualisointi**

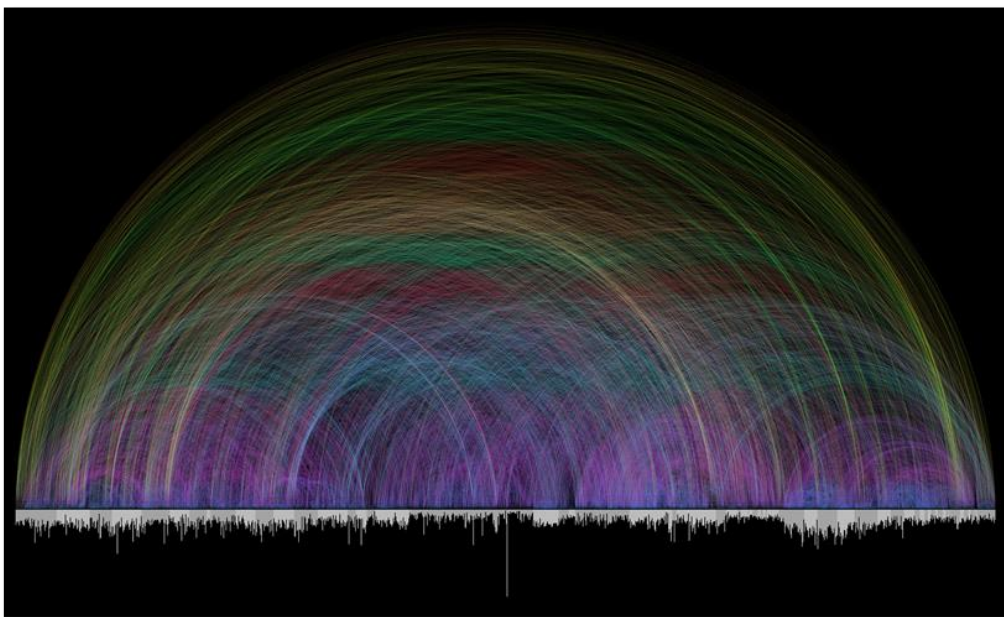
Yhä kasvavien datamäärien edessä, tiedon kuvaaminen ja esittäminen muodostuu alati suuremmaksi ongelmaksi. Miten esittää tulokset ja päätökset selkeämmin? Useat mediat ja alan kolumneja kirjoittavat henkilöt ovatkin esittäneet datan visualisoinnin yhtenä Big Dataa kuvaavista V-määreistä ja pitävät sitä avaintekijänä Big Datan -aikakaudella [McNulty 2014; Khan & Khan 2011]. Visualisoinnin ongelmaa lähestyttäessä, suurimmiksi rajoittaviksi tekijöiksi nousevat ihmisen rajoittuneet kognitiiviset kyvyt käsitellä laajoja visualisointeja sekä visualisoinnin rajautuminen näyttöjen koon mukaan. Monien mielestä visualisoinnin tehokas käyttö onkin ainoa tapa saada Big Datasta sen tuoma arvo esille ja tuottaa se kaikkien käyttöön. [Wang *et al.*, 2015; NGrain 2013]. Myös Intelin [2013] raportin mukaan jokaisen yrityksen, joka haluaa saada lisäarvoa Big Datasta, tulisi kääntää katseensa ensisijaisesti datan visualisointia edistäviin työkaluihin. Aikaisemmassa luvussa esiteltyjen kolmen V:n sekä viiden V:n mallien pohjalta, Intel esittääkin näitä tekijöitä yhdistävää mallia (Kuva 2), jossa tiedon visualisoinnilla on merkittävä rooli osana arvonluontiprosessia. Visualisoimalla tieto ymmärrettävään muotoon on arvoa luovien päätösten ja huomioiden tekeminen datan pohjalta mahdollista kaikkien työntekijöiden osalta. Big Datan onnistuneen visualisoinnin merkitys ja visualisoinnin mahdollistaminen toimivatkin päätekijöinä tässä tutkielmassa.





Kuva 2. Intelin esittämä neljän V:n malli korostaen datan visualisointia [Intel, 2013].

Rijmenamin [2016] artikkelin mukaan ensimmäinen Big Dataa vaikuttavasti visualisoiva projekti oli Harrisonin ja Römhildin vuonna 2007 luoma visuaalinen kuvaus (Kuva 3) raamatun 63 779 sisäisestä ristiviittauksesta jaettuna X-akselilla raamatun eri lukujen mukaisesti. Esimerkissä sovelletun datajoukon kokoa voidaan pitää maltillisena Big Datan nykynäkökulmasta katsottuna sisällön rajoituessa hyvinkin alle sataan tuhanteen tietoalkioon. Harrisonin ja Römhildin esimerkistä voidaan myös todeta, että tässä käytetyn datajoukon sisältö oli hyvin tunnettu, jolloin joukon sisäiset tuntemattomat tekijät eivät pääse häiritsemään luotettavan visualisoinnin tekemistä. He kuitenkin antoivat työllään suuntaa sille, miten vaikuttavia visualisointeja voidaan luoda normaalia suurempien tietojoukkojen pohjalta.



Kuva 3. Harrisonin ja Römhildin [2007] tekemä visualisointi raamatun tekstien sisäisistä viittauksista.

Rijmenami [2016] myös toteaa artikkelissaan, että visualisointi ei teknologisesti olisi vaikein prosessillinen osuus Big Datan hyödyntämisessä, mutta se on haasteellisin kokonaisuuden onnistumisen kannalta. Kompleksisen kvantitatiivista sekä kvalitatiivista sisältöä hyödyntävän tarinan esittäminen pelkästään graafien avulla vaatii uudenlaista ja moniulotteisempaa näkökulmaa asiaan.

Big Datan lähteiden suoraviivainen visualisoiminen ei useinkaan ole mahdollista tai tehokasta ilman ennakkoanalyysia [Wang *et al.*, 2015]. IBM:n mukaan parhain lopputulos saavutetaan, kun liiketoiminnan analytiikan ratkaisut yhdistetään osaksi visualisointia [Keahey, 2013]. Yksinkertaisuudessaan tämä IBM:n kohdalla tarkoittaa *Rapidly Adaptive Visualization Enginen* (RAVE) hyödyntämistä valmiin data joukon visualisoimisessa, johon käyttäjä on jo valmiiksi tehnyt datan rakenteen määrittelyn. Tällöin ei voida sanoa, että nämä edustaisivat aikaisemmasta poikkeavia ratkaisuja muuten kuin, että järjestelmä pystyy käsittelemään isompia datamääriä. Käytännössä järjestelmälle tulee yhä määrittää spesifisesti sille annetun tiedon sisältö, jolloin järjestelmä ei pysty itsenäisesti mukauttamaan visualisointia tiedon sisällön tai rakenteen muuttuessa. Tiedon ja Big Datan visualisointiin liittyviä tekijöitä tullaan käsittelemään vielä tarkemmin tutkimuksen luvussa 3.

### 2.3 Big datan hyödyt ja ongelmat

Yritykset ja valtioiden toimijat näkevät Big Datassa suuren mahdollisuuden. Yhdysvaltain Valkoinen talo esitti jo vuonna 2012 julkaisemassaan tukirahoituskampanjassaan digitaalisen tiedon hyödyntämisen mahdollisuutena ratkaista valtion isoimmat ongelmat energiateollisuuteen, terveydenhuoltoon ja maan puolustukseen liittyen [OSTP, 2012]. Valkoisen talon tukirahoituskampanjassa tuodaan esille rahalliset avustukset visuaalisten ja graafisten tekniikoiden toteuttamiseen ja niiden kehittämiseen kompleksista dataa varten. Erityisesti tarve nopeasti reagoivaan ja päätöksentekoa helpottavaan visualisointityökaluun tuodaan esille erilaisissa maan puolustukseen liittyvissä skenaarioissa ja järjestelmätarpeissa.

Big Datan hyödyntämisen on todettu tarjoavan paremman yleiskuvan yrityksen tuotantoprosessista (*supply chain*) ja johtavan asiakkaiden parempaan ymmärtämiseen. Big Dataa on hyödynnetty uusien innovaatioiden löytämiseen [Monino, 2016] ja tuotantoprosessien optimoimiseen [Klein & Gorton, 2015]. Yritysten jotka hyödyntävät Big Dataa, on arvioitu saavan merkittävää etumatkaa kilpailijoihin nähden heidän pystyessä ennakoimaan markkinan toimintaa tehokkaammin. Rajaraman [2016] antaa esimerkin laajasta kahvilaketjusta, jolla on useita liikkeitä ympäri Yhdysvaltoja. Kahvila lanseerasi uuden kahvisekoituksen ja turvautui sosiaalisen median kautta

kerättyyn Big Dataan analysoidessaan kuluttajakertomuksia uuden kahvilaadun suhteen. Julkistuspäivän iltapäivään mennessä yritys oli pystynyt seulomaan suurimmaksi ongelmaksi korkean hinnan ja muuttaneet tätä dynaamisesti ensimmäisen myyntipäivän aikana. Reaaliaikaisessa seurannassa negatiiviset palautteet hinnasta olivat loppuneet. Voidaankin todeta, että kun yritys saa prosessinsa sekä järjestelmänsä reaaliaikaista (Velocity) Big Dataa tukevaan tilaan, ja yritys pystyy tuottamaan kerätystä tiedosta merkityksellistä arvoa, yrityksen tehokkuus siirtyy seuraavalle tasolle.

Big Datan käyttöön liittyvät ongelmakohdat voidaan Akerkarin [2013] mukaan jakaa kolmeen ryhmään: Datan kompleksisuuteen (määrä, laajuus, eroavaisuus), datan prosessoimiseen (yhtäläisyyksien löytäminen, datan muokkaaminen ja analysoiminen) sekä datan hallintaan (yksityisyys, turvallisuus, eettisyys). Rajaramankin [2016] nostaa tutkimuksessaan erityisesti ylös tilanteen datan hallinnasta: ”*Big Data on kuin kaksiteräinen miekka. Samalla kun se tarjoaa paljon hyödyllistä informaatiota kansalaisille, se johtaa myös yksityisyyden häviämiseen*”. Käyttäjät ovat eri järjestelmiin tietoja syöttäessään harvoin tietoisia siitä, että kuinka pitkälle annettua tietoa voidaan yhä jalostaa ja mitä kaikkea järjestelmät seuraavat käyttäjän toiminnassa. Käyttäjien seuraaminen verkkopalveluissa on sosiaalisen median palveluiden kautta luotu yhä helpommaksi. IP-osoitteiden, sosiaalisen median käyttäjätunnusten (joita yhä useammin käytetään palveluissa vaihtoehtoisena kirjautumistapana) ja laitteiden tunnistetietojen yhdistelyiden avulla eri käyttäjien ja eri palveluiden tietoja voidaan yhdistellä suuremmaksi kokonaisuudeksi ja käyttäjän toiminnalliseksi kartastoksi. Erityisen huolestuttavan tilanteesta tekee se, että tietomurtojen kautta ihmisistä vuotaa yhä kasvavassa määrin erilaista dataa väärin käsiin. Esimerkkinä Yahoo-palvelun murtaminen [Yahoo, 2016], mitä on tähän mennessä pidetty yhtenä isoimpana tunnetuista tietomurroista. Tietomurtojen lisäksi on tullut ilmi huolestuttavia tapauksia, joissa myös isot käyttäjäpalvelut ovat hyödyntäneet ihmisten henkilökohtaisia tietoja poliittisiin tai kaupallisiin tarkoituksiin. Näistä tunnetuimpana on ainakin 72 miljoonan Facebook – käyttäjän tietojen luovuttaminen Cambridge Analyticalle Facebookin luvalla [Facebook, 2018].

Rajaraman [2016] yhä lisää, että datan sisältöön ei tule sokeasti luottaa vielä datan analysoinnin jälkeenkään. Big Dataa käsiteltäessä ei voida turvautua siihen, että kun tietoa on kerätty massiivisesti data ikään kuin automaattisesti paljastaisi sisältönsä ja sisäiset suhteensa. Tämä on Rajaramanin mukaan kaukana todellisuudesta, sillä harvoin on mahdollista saada käyttöönsä oikeasti kaikkea dataa. Rajaraman mainitsee esimerkkinä vaalien ennakkotulosten arvioinnin, sillä näissä otanta on pieni, jolloin Big Dataan pohjautuvaa ennakoivaa (*predictive*) analyysia ei voi luotettavasti tehdä. Toisena ongelmana nähdään dataan ja tuloksiin luottaminen sokeasti data-analyysin tekemisen jälkeen. Mahdollisuutena on, että analysointivaiheessa datan sisäiset korrelaatiot ja suhteet on muodostettu tai ymmärretty väärin, jolloin analyysiprosessin tulos on virheellinen.

Big Datan tietokantahallintajärjestelmän (Big Data Management System, BDMS) voidaan pitää yleisnimityksenä Big Datan hyödyntämisessä käytetyille järjestelmäratkaisulle. Järjestelmän tulee kyetä säilöämään ja prosessoimaan petatavun kokoisia datajoukkoja ja niiden tulee olla arkkitehtuuriltaan skaalautuvia, jaoteltuja, tehokkaita sekä virhesietoisia [Fernandez *et al.*, 2014; Marcos *et al.*, 2013]. Jatkuvan datamäärän kasvun vuoksi Big Data -pohjaisten järjestelmien haasteena on erityisesti skaalautuvuus. Järjestelmän tulee pystyä toimimaan myös pitkälle tulevaisuuteen, vaikka tietosisältö olisi laajentunut moninkertaiseksi lähtötilanteesta. Tämän lisäksi reaaliajassa toimivien järjestelmien tulee mukautua virheisiin, jossa esimerkiksi hajautetussa laskennassa käytetyistä koneista osa lopettaa toimintansa tai tietovarastoina käytetyt moduulit täyttyvät [Neves & Bernardino, 2015]. Myös Klein ja Gorton [2015] alleviivaavat skaalautuvuuden ongelmaa ja toteavat, että Big Data järjestelmän käyttöönoton jälkeen järjestelmään tulevien syötelähteiden määrän ja kasvun kontrolli häviää tyystin. Täten järjestelmälle ei voi etukäteen asettaa minkäänlaisia määrytyksiä sen laajuudesta, sillä nämä voivat tulla nopeasti vastaan. Siksi järjestelmän tilaa tulisi jatkuvasti monitoroida ja seurata järjestelmän sisäisiä trendejä. Tällöin kattavan seurannan kautta pystytään ennakkoon vaikuttamaan järjestelmässä tapahtuviin muutoksiin tekemällä muutoksia hajautetussa järjestelmä- ja palvelinympäristöissä [Klein & Gorton, 2015]. Tarkemmin Big Data järjestelmien ongelmiin palataan vielä luvussa 3, jossa käydään läpi erityisesti virtuaalitodellisuutta hyödyntävien järjestelmien luomat haasteet.

On arvioitu, että vuonna 2020 esineiden internet (*Internet of Things, IoT*) tulee yksistään tuottamaan 4 zettatavun verran dataa vuoden aikana [Turner *et al.*, 2014]. Tällöin syötelaiteiden ja niiden luoma tietomäärä asettaa yhä kasvavan ongelman Big Data järjestelmien skaalautuvuudelle. Hajautettuun verkkoon tulee lisätä dynaamisesti uusia prosessoivia laitteita ja datan käyttö tulee jakaa alustalle tasaisesti. Watsonin [2014] mukaan on vaikea luoda yhtä geneeristä alustaa Big Dataa varten, jonka organisaatiot voisivat helposti ja nopeasti ottaa käyttöönsä. Usein tämä vaatiikin laajempaa järjestelmien kustomointia riippuen järjestelmän ympäristöstä ja järjestelmään ajettun tiedon ominaisuuksista. Big Datan trendiin ja käsitteeseen kuuluu vahvasti myös pilvipalvelut, jotka ovat alkaneet voimakkaasti yleistyä, mikä avaa tehokkaampia mahdollisuuksia isojen data määrien hallitsemiseen [Teräs & Raghunathan, 2015]. Pilvipalveluiden myötä tiedot siirtyvät käyttäjien lokaaleilta kovalevyiltä pilvipalveluiden servereille, jolloin saatavilla olevan tiedon määrä kasvaa ja tiedon käyttö helpottuu.

### 3. TIEDON VISUALISOIMINEN

Tiedon visualisoinnin kappaleessa annetaan aluksi yleiskuvaus tiedon visualisoinnin alasta. Tiedon visualisoinnin perusta ja määritelmä käydään lävitse. Tämän jälkeen kuvataan visualisoinnin käyttötapauksia ja mahdollisuuksia esimerkkien avulla. Määrittelyiden jälkeen esitellään visualisoinnin tekniikoita ja menetelmiä. Näitä menetelmiä arvioidaan erityisesti virtuaalitodellisuuden ja Big Datan käyttöön soveltuvuuden näkökulmista. Lopuksi tehdään yhteenveto virtuaalitodellisuuteen ja Big Datan visualisointiin parhaiten soveltuvista tekniikoista.

#### 3.1 Visualisointi yleisesti

Visualisoinnilla tarkoitetaan jonkin asian tekemistä havainnoitavaksi näköaistin avulla. Erityisen tärkeää visualisoinnissa on muistaa ihmisten sisäisten mallien muodostumisen tukeminen. Donalekin ja kumppaneiden [2014] mukaan visualisointi on pääväylä kvantitatiivisen tietosisällön ja ihmisen tietoisuuden välillä. Ajatuksen pohjana on, että ihminen ei kykene täysin ymmärtämään asioita, joita ihminen ei pysty jollakin tavalla visualisoimaan. Ihmisille onkin kehittynyt merkittävä taito kaavojen ja yhteyksien tunnistamisessa. Tästä syystä taito löytää tietoa datan ohjaamassa (*data-driven*) tieteessä pohjautuu merkittävästi onnistuneeseen datan visualisointiin, jossa datan tutkiminen olisi tehokasta ja joustavaa.

Visualisointi nähdään monialaisena käsitteenä sisältäen tekniikoita tietokonegrafiikasta, kuvankäsittelystä, konenäöstä, tietokoneavusteisesta opetuksesta, geometrisesta mallinnuksesta, psykologiasta ja käyttöliittymäsuunnittelusta [Haber & McNabb, 1990]. Visualisoinnin onkin nähty edistäneen kehitystä monilla eri tutkimusaloilla [Johnson *et al.*, 2005]. Visualisoinnin ala jaetaan perinteisesti kahteen eri osa-alueeseen: Tieteelliseen visualisointiin (*scientific visualization*) ja tiedon visualisointiin (*information visualization*).

Tieteellisessä visualisoinnissa pyritään yhdistämään ihmisen kognitio osaksi tietokoneita ja niiden tuottamaa grafiikkaa. Käytännössä tieteellisen visualisoinnin ala keskittyy tieteellisten tutkimustulosten liittämiseen osaksi reaali maailman prosesseja. Tutkimus alana tieteellisessä visualisoinnissa pyritään helpottamaan tiedon ymmärtämisen prosessia ohjelmistotyökalujen avulla, jotka tarjoavat staattisia tai interaktiivisia visualisoinnin esityksiä. [Johnson *et al.*, 2005].

Tiedon visualisointi syntyi omana haaranaan ihmisen ja tietokoneen välisen vuorovaikutuksen tutkimusalan (*Human Computer Interaction, HCI*) syntyessä 1980 -luvun lopulla. Tiedon visualisoinnin tutkimusalan tarkoituksena on tiedon ymmärtämisen ja käsittelyn

helpottaminen, mutta pääpaino tutkimuksessa on erityisesti mentaalimallien ja tietokonegrafiikan hyödyntämisessä.

Brown ja kumppanit [1996] näkevät, että visualisoinnilla on kolme keskeistä päämäärää: tutkiminen, analysointi ja esittäminen. Visualisointia siis käytetään olemassa olevan datajoukon tarkasteluun ja mielenkiintoisten rakenteiden hahmottamiseen jatkoanalysoimista varten. Visualisointia voidaan yhä käyttää hypoteesien ja saatujen tulosten varmentamiseen ja esittämiseen esimerkiksi muille kollegoille tai julkiselle yleisölle. Visualisointi on kuitenkin harvoin eriytetty, itsenäinen prosessinsa: se on usein tarpeellinen, mutta ei yksistään riittävä tapa ongelmien ratkaisemiseen. Tiedon visualisoiminen vaatiikin rinnalleen yhä vahvemmin analyttisiä välineitä ja tekniikoita kuten tilastotiedettä, tiedonlouhintaa ja kuvaprosessointia. Johnson ja kumppanit [2005] esittävätkin raportissaan yhtenä alan suurimmista ongelmista visualisointitutkimuksen mukautuvuuden muiden alojen tiedonvisualisointitarpeisiin. Tämän haasteen pohjalta onkin esitelty uusi poikkitieteellinen käsite tieteellisen ja informaation visualisoinnin rinnalle: Visuaalinen analytiikka (*Visual Analytics*). Visuaalinen analytiikka koostaa aikaisempia lähestymistapoja yhteen, mutta painottaa koneoppimista ja mallien rakentamista tiedon perusteella. Luotuja malleja usein tarkennetaan ja kehitetään visualisoinnin avulla.

Tässä tutkielmassa visuaalisuutta käydään läpi erityisesti sen analyttisen puolen osalta, sillä tutkielmassa pyritään kartoittamaan ratkaisuja datan helpompaan analysoimiseen hyödyntämällä virtuaalitodellisuuden työkaluja. Tiedon visualisointiin pohjautuen tutkielmassa pyritään löytämään käyttäjien mentaalimalleja tukevia ja virtuaalitodellisuuden mahdollisuuksia hyödyntäviä visualisointiratkaisuja Big Dataksi luokiteltavan tiedon esittämistä varten. Tutkielmassa virtuaalitodellisuuden teknologia nähdään uutena työkaluna ja sitä hyödyntämällä luotuja visualisointeja pyritään vertaamaan perinteisessä työasemaympäristössä toteutettujen visualisointien suhteen. Tutkielman ensisijaisena päämääränä on selvittää lisääkö virtuaalitodellisuus tiedon visualisoinnin tehokkuutta ja ymmärrystä verrattuna perinteiseen työasemaympäristöön ja 2D-visualisointeihin.

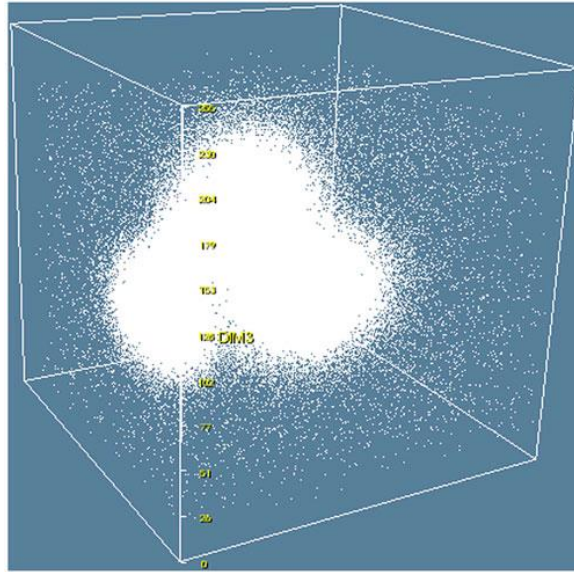
### **3.2 Visualisoinnin tyypit**

Tiedon visualisoinnin menetelmät ovat digitalisaation myötä edenneet vauhdilla eteenpäin viime vuosikymmenien aikana. Alun perin tiedon visualisointia hyödynnettiin karttatieteissä ja liiketoimintaan pohjautuvan tilastollisen tiedon esittämisessä [Few & Edge, 2007]. Nykykäsitteen mukaisen tiedon visualisoinnin katsotaan toteutuneen ensimmäisen kerran William Playfairin teoksessa *Commercial and Political Atlas* [1786], jossa hän esitteli Englannin kaupankäyntiin



pohjautuvaa tilastoa graafien avulla. 2000-luvun aikana tiedon visualisoinnissa on keskitytty erityisesti edistämään tiedon sisällön interaktiivisuutta ja erilaisia vuorovaikutustapoja visualisoituun sisältöön liittyen [Kehrer *et al.*, 2012]. Interaktiivisuuden merkitys korostuu, kun visualisointi toteutetaan 3D-mallinnettuna, jolloin vuorovaikutuksessa hyödynnettävät mahdollisuudet kasvavat dimensioiden lisääntyessä. Nykyään tiedon visualisoinnissa ongelmia tuottaa datan suuri määrä ja visualisoinnin skaalautuvuus datan määrän mukaisesti, joka määriteltiinkin tutkielman kappaleessa 2.1. Tämän mukaisesti skaalautuminen on määritelty yhdeksi isoimmista ongelmista Big Datan hyödyntämisessä. Lisäksi Chen & Zhang [2014] toteavat, että oikean visualisointitavan valinta on merkittävin tekijä Big Datan visualisointiprosessissa.

Perinteisimmiksi tiedon visualisointitavoiksi voidaan luokitella: taulukot, piirakkakaaviot, pylväsdiagrammit, viivakaaviot, pisteparvet, vuokaaviot ja aikajanat. Perinteiset visualisointimenetelmät muuttuvat kuitenkin tehottomiksi ja epäselviksi datan määrän kasvaessa. Esimerkiksi piirakkakaaviossa hyödynnetään värikoodausta eri osien erottamiseen toisistaan. Big Datan kohdalla tietueiden määrä on kuitenkin niin laaja, että toisistaan erottuvat värikoodaukset loppuvat kesken. Myös Ware [2004] toteaa kirjassaan, että ihmisen visuaalinen työmuisti on rajoittunut kolmesta viiteen kohteeseen kerralla, jolloin värikoodausta tulisi hyödyntää vain pienissä visualisoinneissa. Yksi suosituimmista monimuuttuja-analyysissä käytetyistä visualisointitekniikoista on pisteparvitaulukko *Scatter Plot Matrix (SPLOM)* [Card *et al.*, 1990]. Pisteparvet auttavat segmenttien, raja-arvojen, trendien ja korrelaatioiden löytämisessä, mutta laajaa moniulotteista dataa kuvattaessa joudutaan hyödyntämään useita eri arvoja kuvaavia pisteparvitaulukkoita. Toinen ongelma pisteparven kohdalla esiintyy, kun datapisteitä on niin paljon, että pisteet alkavat piirtymään toistensa päälle (*overplotting*). Visualisoinnin sotkeutuminen (*clutter*, *visual clotting*) on päälle piirtymisestä juontuva ongelma, jolloin ei voida enää hahmottaa datapisteiden sijaintia ja raja-alueita visualisoinnin sisällä [Du *et al.*, 2016; Fisher, 2016]. Visuaalisuuden heikkenemisen ongelma on esitettyä kuvassa 3.



Kuva 3. Visuaalisuuden heikkenemisen ongelma (visual clotting) kuvattuna 3D-pisteparvessa.  
[Du et al., 2016].

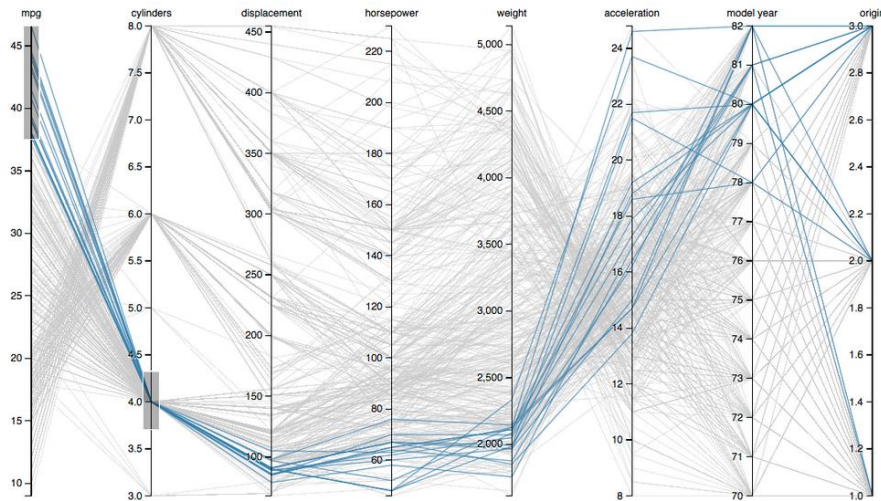
Big Dataa kuvattaessa tarpeet visualisoinnille säilyvät hyvin samanlaisina kuin normaaleissakin tilanteissa, mutta näiden tarpeiden saavuttaminen vaatii normaalia enemmän työtä. Olshannikova ja kumppanit [2015] määrittävät tutkimuksessaan Big Datan visualisoinnin merkittävimiksi tekijöiksi seuraavat: piilotettujen yhteyksien tai anomalioiden identifiointi, spesifisten arvojen etsimisen joustavuus, eri arvojen kvantitatiivinen vertailu ja käyttäjän reaaliaikainen vuorovaikutus visualisoinnin kanssa. Näiden tavoitteiden saavuttaminen kuitenkin vaikeutuu käsiteltäessä huomattavasti isompia datajoukkoja, joihin aikaisemmin esitetyt perinteiset visualisointitekniikat taipuvat huonosti. Perinteisempien visualisointitekniikoiden heikon Big Data soveltuvuuden takia seuraavaksi käydään lävitse normaalista poikkeavampia visualisointitapoja, jotka tukevat kompleksisen ja monia ulottuvuuksia sisältävän datan käyttöä. Visualisointitekniikoista arvioidaan niiden soveltuvuutta isojen tietomäärien visualisoimiseksi ja esitellään tekniikoiden vahvuudet ja heikkoudet.

### 3.2.1 Parallel coordinates

*Rinnakkaiskoordinaattien* tekniikkaa käytetään yksittäisen dataelementin piirtämiseen useiden dimensioiden välillä. Täten tekniikka soveltuu erityisen hyvin moniulotteisen datan esittämiseen ja tekniikkaa on laajasti hyödynnetty [Inselberg & Dimsdale, 1990]. Visualisoinnin pääperiaatteena on esittää jokaisen tietueen arvot sarjana vierekkäisiä akseleita ja jokainen arvo on linkitetty osaksi omaa akseliaan. Visualisointityyli voidaan nähdä yksinkertaisemmillaan taulukkona, jonka rivien sarakkeiden välille on vedetty yhdistävät viivat. Tämän takia rinnakkaisten koordinaattien avulla



voidaan visualisoida useita arvoja sisältävät tietueet yhdessä kuvaajassa. Myös rinnakkaisten koordinaattien visualisointi sisältää saman ongelman kuin pisteparvi: Tietomäärän kasvaessa viivat alkavat vahvasti piirtymään toistensa päälle, jolloin visualisoinnista on enää vaikea erottaa yksittäisiä kohteita. Täten visualisointitekniikkaa on vaikea hyödyntää isoille data joukolle.



*Kuva 4. D3-kirjaston luoma esimerkki rinnakkaisten koordinaattien visualisoinnista, jossa on kuvattuna automallien tietoja.*

### 3.2.2 Star coordinates

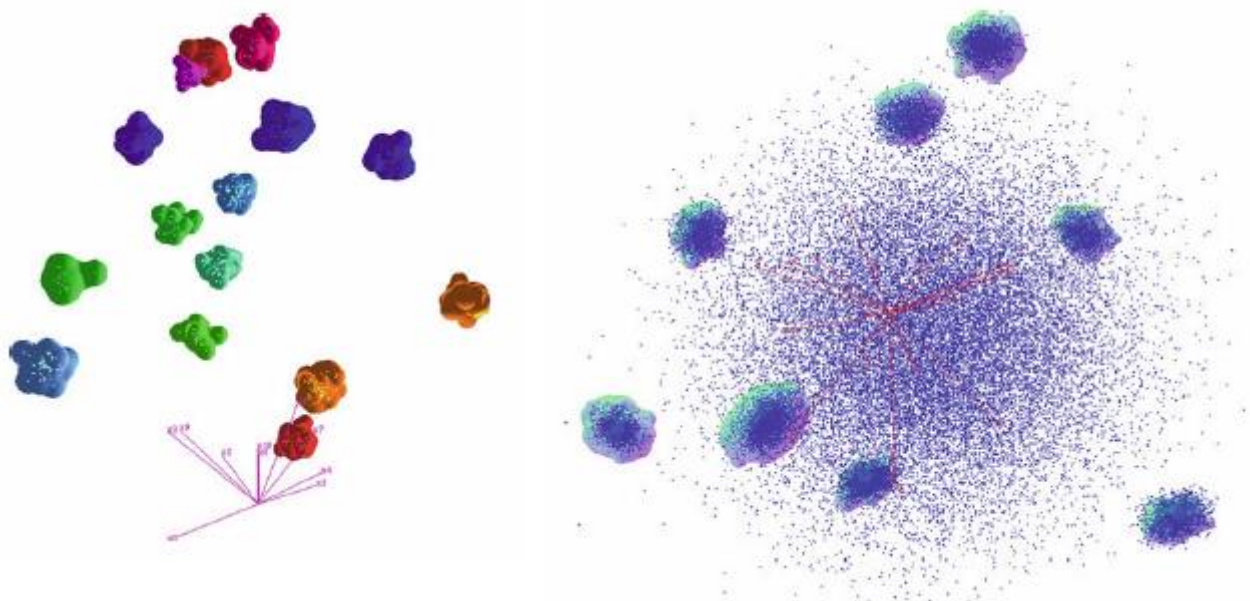
Tähtikoordinaatit ovat yksinkertainen, tehokas ja hyvin tunnettu tiedon interaktiotapa moniulotteisen datan visualisoimiseen. Yleisimmin tekniikkaa käytetään tutkimuksellisiin tarkoituksiin kuten klusterien analysoimiseen, poikkeavien havaintojen suodattamiseen tai trendien havaitsemiseen [Rubio-Sanchez & Sanchez, 2014]. Tähtikoordinaattien käytön tarkoituksena onkin luoda helposti ymmärrettäviä moniulotteisia visualisointeja, jotka tukevat datajoukon sisällön hahmottamisprosessia. Täten päämäärä ei ole numeerinen analyysi vaan yleiskäsityksen luominen tiedon sisällöstä [Kandogan, 2000; 2001].

Tähtikoordinaattitekniikan ideana on järjestää koordinaattiakselit ympyrän sisälle, jossa jokainen samanpituinen akseli omaa lähtökohtaisesti yhtä suuren kulman ympyrän keskustan ja akselin välillä. Chenin [2014] mukaan tähtikoordinaattitekniikkaa voidaan hyödyntää jopa miljardien eri tietueiden yhtä aikaiseen kuvaamiseen, sillä visualisointi tukee päällekkäin menevien tietojen aggregointia syvyystiedon muodossa toisin kuin esimerkiksi rinnakkaiset koordinaatit. Lisäksi Chen toteaa tutkimuksessaan, että tähtikoordinaatit soveltuvat erityisen hyvin Big Datan visualisointiin, sillä visualisoinnissa ei tarvitse laskea tietueparien (*pairwise*) välisiä etäisyyksiä, vaan tieto

etäisyyksistä säilötään visualisoinnin pohjalla toimivaan malliin. Tämä erottelu mahdollistaa esimerkiksi aikaisemmin mainitun syvyystietojen hyödyntämisen visualisoinnissa.

Tähtikoordinaatteihin pohjautuvaa klusterianalyysia ja validointia on hyödynnetty esimerkiksi Kandoganin tutkimuksessa [2001], VISTA -järjestelmässä [Chen & Liu, 2004] sekä Longin & Linsen [2011] moniulotteista dataa käsittelevässä tutkimuksessa. Tähtikoordinaattien hyödyntäminen vaatii myös käyttäjän interaktiota ja harvoin visualisointi on heti alussa käyttäjälle arvoa tuottava ilman käyttäjän tekemiä päätöksiä ja analyysiä. Tärkeimmäksi näistä voidaan lukea elementin painoarvon mukauttamisen visualisoinnin sisällä. Arvon muuttaminen tapahtuu skaalaamalla elementtien akselien pituutta visualisoinnissa. Muihin yleisimpiin vuorovaikutustapoihin lukeutuvat esimerkiksi arvoalueiden antaminen, tietueiden valitseminen ja arvojen välisten korrelaatiopainotusten muuttaminen [Kandogan 2000]. Muokkauksen jälkeen visualisointi joudutaan piirtämään kokonaan uudelleen. Täten käyttäjän tekemien uudelleenmääritysten takia vuorovaikutusprosesseihin tulee kiinnittää erityistä huomiota Big Dataa hyödyntävissä järjestelmissä, jotta muutokset pystytään prosessoimaan nopeasti käyttökokemusta heikentämättä.

Tähtikoordinaattien tukiessa moniulotteista dataa ja sen soveltuessa myös massiivisten tietojoukkojen esittämiseen tähtikoordinaatit voidaan nähdä kattavimpana visualisointivaihtoehtona Big Dataa varten, sillä teoriassa datajoukon suuruudella ei ole vaikutusta visualisoinnin onnistumiseen. Lisäksi Kandoganin [2000; 2001] tutkimusten mukaan, tähtikoordinaattien tekniikka soveltuu erityisesti data-analyysin ensimmäisiin vaiheisiin, joissa pyritään hahmottamaan tietojoukon yleisiä ominaisuuksia. Tämä prosessinvaihe tulee olemaan painotettuna myös tämän tutkielman visualisointijärjestelmän kehityksessä.

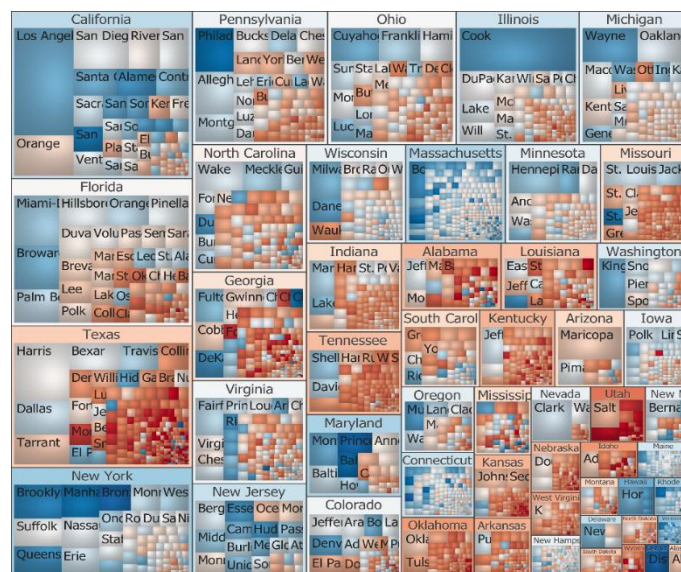


*Kuva 5. Long & Linsen [2011] kuvasivat 10-ulotteista (vas.) ja 20-ulotteista dataa 3D-muodossa tähtikoordinaattien avulla.*

### 3.2.3 Tree map

Klassinen puukartta tarjoaa mahdollisuuden aggregoida tietoa datajoukon hierarkian ylemmillä tasoilla ja tarjoaa hierarkian avulla yhä tarkempia näkymiä tiedon sisällöstä [Shneiderman, 1992]. Puukartoilla voidaan esittää hierarkkista dataa jaoteltuna sisäkkäisiin suorakulmioihin. Jokainen tietoelementti toimii puun yhtenä haarana, jolle annetaan suorakulmio, mikä sisältää elementin sisältämät arvot jaettuna uusiin ja pienempiin suorakulmioihin. Puukartoissa hyödynnetään tiilien kokoja ja värien korrelaatiota, jolloin tiedon sisältö ja merkittävyys on käyttäjälle helpommin hahmotettavissa. Puukarttojen hyvänä puolena on se, että visualisointi käyttää tehokkaasti tilaa hyödykseen ja mukautuu pienempäänkin tilaan.

Puukarttojen sisältäessä hyviä ominaisuuksia, se voitaisiin sovellettuna versiona nähdä hyvänä vaihtoehtona tiedon visualisoinnille virtuaalitodellisuuden ympäristössä. Kuitenkin erityisesti Puukarttojen kohdalla visualisointitekniikan hyödyntäminen riippuu yhä vahvemmin datan sisällöstä, sillä kuten mainittua puukarttojen visualisointi nojaa vahvasti tiedon hierarkkisuuteen. Tietojoukon ollessa sisällöltään ja ulottuvuuksiensa osalta vahvasti heterogeenistä tietojoukon sisäisistä korrelaatioista tulee epäluotettavia ja puukartaston luomasta visualisoinnista tulee hankalasti tulkittava.



*Kuva 6. Puukartta luotuna USA:n vuoden 2012 presidentinvaalien äänestystuloksen jakautumisesta.*

### 3.3 Visualisoinnin työkalut

Datan visualisointi tarkoittaa tiedon esittämistä järjestelmällisessä muodossa sisältäen muuttujien ja yksikköjen tiedot [Khan & Khan, 2011].

Edut	Prosentit (%)
Parantunut päätöksenteko	77
Parempi ad-hoc data-analyysi	43
Parantunut yhteistyö ja tiedon jakaminen	41
Itsepalvelumahdollisuuksien tarjoaminen loppukäyttäjille	36
Kasvanut ROI (Return on investment)	34
Aikasäästöt	20
Vähentynyt IT-kuorma	15

*Taulukko 1. Datan visualisointityökalujen hyödyt [Sucharitha et al., 2014].*

Sucharitha ja kumppanit [2014] esittivät Big Dataa käsittelevässä kyselytutkimuksessaan visualisointityökalujen suurimmat hyödyt (Taulukko 1). Tutkimuksen tulosten pohjalta suurin osa vastaajista oli sitä mieltä, että tiedon visualisointia voidaan erityisesti käyttää osana tehokkaampaa päätöksentekoprosessia. Visualisointityökalujen käyttö tuo tiedon sisällön yhä laajemman käyttäjäsegmentin hyödynnettäväksi, kun tieto on muunnettu ymmärrettävään ja helpommin lähestyttävään muotoon. Tällöin visualisoinnit tarjoavat yrityksille mahdollisuuden havaita tuotteiden, myynnin ja asiakkaiden välisiä korrelaatioita, jolloin kohdemarkkinointi tehostuu. Visualisointi tarjoaa myös työvälineen yrityksen toiminnan seurantaan, jota voidaan hyödyntää riskianalyysien tekemisessä.

Datamäärien ja datan kompleksisuuden kasvaessa yhä pidemmälle menevien tavoitteiden saavuttaminen työkalujen avulla vaikeutuu. Isoja tietojoukkoja käsiteltäessä yleensä tukeudutaan tiedon tiivistämiseen käsittelyn helpottamiseksi, mutta Big Dataa tiivistettäessä ei voida suoraan määrittää, mikä osa tiedosta voidaan jättää pois [Olshannikova et al., 2015]. Visualisointityökalujen haasteena onkin yhdistää abstrakti tieto osaksi reaali maailmaa visuaalisen esityksen kautta. Näistä tekijöistä johtuen visualisointityökalujen tulisi jatkossa tarjota yhä joustavampia ratkaisuita datajoukon määrittelyyn ja analyysiin. Tämän lisäksi visualisointityökalujen esityksen tulisi myös täyttää ainakin seuraavat kolme vaatimusta: Ilmaisevuus (esitä vain tieto, jota

data sisältää), tehokkuus (pohjautuen ihmisen kognitiiviseen havainnointiin) ja soveltuvuus (visualisoinnin hyötykustannus suhde) [Miksch & Aigner, 2014; Muller & Schumann, 2003].

Datajoukkojen alati laajentuessa ja muuttuessa yhä kompleksisemmiksi tutkijat ovat alkaneet painottaa vuorovaikutustyökalujen tärkeyttä visualisoinneissa. Tällöin päätös sisällön rajaamisesta, tulkinnasta ja joukon sisäisten yhteyksien etsinnästä annetaan yhä vahvemmin käyttäjälle. Päämääränä on tarjota mahdollisimman paljon dataa käyttäjän nähtäville ja tarjota mahdollisimman tehokkaat työkalut datan tulkintaan. Tällöin visualisoinnin käytöstä tulee joustavampaa, minkä tulisi tehostaa analysoinnin tekemistä. Wang ja kumppanit [2015] toteavatkin tutkimuksessaan, että käyttäjän osallistaminen ja interaktiivisten työkalujen käyttö on hyvin tärkeää, sillä staattiset visualisoinnit eivät tee autuaaksi ja ovat huomattavasti tehottomampia datajoukkoa analysoitaessa. Visualisoinneissa tulisikin nojautua ihmisen kognitiiviseen kykyyn havaita visuaalisia malleja ja siirtää kriittinen ajattelu käyttäjän vastuulle. Khan & Khan [2011] nimeävät tutkimuksessaan vuorovaikutteisen visualisoinnin vaiheet:

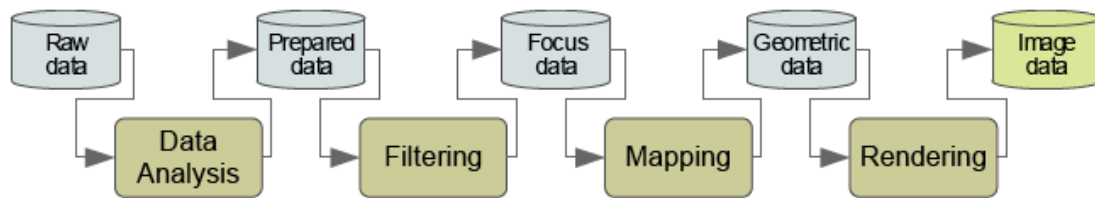
1. *Valitseminen:* Käyttäjällä tulee olla mahdollisuus valita yksittäinen tietue, osajoukko tai koko datajoukko oman kiinnostuksen mukaisesti.
2. *Linkitysten tekeminen:* Mahdollisuus linkittää tietueita toisiinsa ja vertailla niiden sisältämiä arvoja eri näkymissä.
3. *Suodattaminen:* Auttaa käyttäjää muuttamaan esillä olevan tiedon määrää ja auttaa keskittämään fokuksen niihin elementteihin, joista ollaan kiinnostuneita.
4. *Uudelleen järjestäminen:* Spatiaalisen näkymän ollessa tärkein tapa visuaalisessa havainnoinnissa tulee käyttäjällä olla mahdollisuus muuttaa tiedon asettelutapaa ja näkymää uusien näkökulmien saamiseksi.

### **3.4 Visualisoinnin prosessi**

Visualisointityökalujen haasteiden ja vaatimusten esittelyn jälkeen käydään läpi yleisluontoinen prosessimalli visualisoinnin toteuttamiseksi. Tietotekninen prosessi raa'an datan johtamisesta visualisoinniksi (*visualization pipeline*) noudattaa vahvasti edellisessä kappaleessa Khan & Khanin [2011] käyttäjän näkökulmasta tekemää vuorovaikutteisen visualisoinnin prosessia. Järjestelmätason visualisoinnin prosessi on määritelty koostuvan seuraavista vaiheista: Tiedon analysoiminen, suodattaminen, kartoittaminen ja kuvantaminen [InfoVis, 2018].



1. *Tiedon analysoinnissa* data valmistellaan visualisoimista varten esimerkiksi poistamalla puuttuvat tai virheelliseksi määritetyt arvot tai suodattamalla osa ei-halutuista arvoista pois.
2. *Suodattamisessa* tietojoukosta valitaan halutut osat visualisointia varten. Valinnat määrittävät usein käyttäjän tekemänä.
3. *Kartoittamisessa* data, josta ollaan kiinnostuneita, liitetään osaksi geometrisiä primitiivejä (esimerkiksi pisteet ja viivat) ja niiden ominaisuuksia (väri, sijainti, koko).
4. *Kuvantamisessa* aikaisempien vaiheiden pohjalta muodostettu geometrinen data muunnetaan kuvalliseen ja visuaaliseen muotoon.



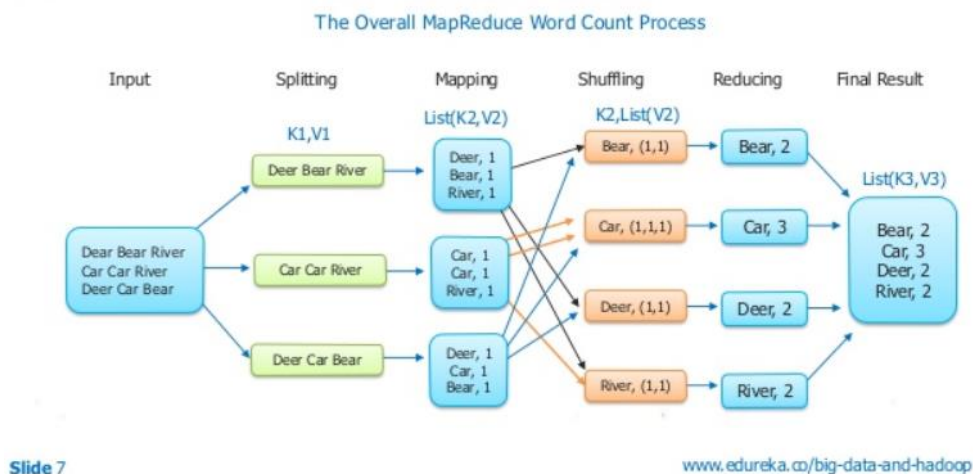
Kuva 7. Visualisoinnin prosessi kuvattuna.

Perinteistä visualisoinnin prosessia hyödyntäviä järjestelmiä on kehitetty useita vuosien varrella. Nykyään visualisoitavan datan määrä asettaa näille kuitenkin haasteensa, sillä tiedon käsittely, suodattaminen ja yhteen liittäminen suoritetaan offline-tilassa paikallisesti välittämättä kustannuksista. Datan määrän kasvaessa ongelmat kasvavat, kun yhä enemmän siirretään tietoa säilövän moduulin (tietokanta) ja kuvantamisen suorittavan moduulin (käyttöliittymä) välillä (*client - server malli*) [Vo et al., 2011]. Myös Moreland [2013] toteaa tutkimuksessaan, että nykyiset ahneet visualisoinnin algoritmit on suunnattu tarjoamaan lyhytaikaista laskentaa isolle datajoukolle. Toisena ongelmana Moreland näkee visualisointijärjestelmien huonon skaalautuvuuden jatkuvaan ja dynaamiseen datajoukkojen käsittelyyn. Aikaisemmat visualisointityökalut ovat muuntautuneet huonosti tukemaan uusia ja kompleksisempia tietorakenteita, joten ratkaisuja on lähdetty hakemaan muualta. Suurten datajoukkojen käsittelyä ja visualisointia varten onkin kasvavassa määrin alettu käyttämään *MapReduce*-ohjelmointimallia. MapReduce-ohjelmointimalli onkin suunniteltu suurten data määrien nopeaa käsittelyä varten.

### 3.5 MapReduce

MapReduce on yleiskäyttöinen ja kevyt ohjelmointimalli, joka on kehitetty erityisesti tiedon rinnakkaiseen prosessointiin hajautetussa järjestelmäympäristössä. MapReduce on osa Apache Hadoopin avoimeen lähdekoodiin pohjautuvaa kirjastoa, jonka tarkoituksena on mahdollistaa isojen tietomassojen käsittely hajauttamalla tiedon prosessointi pilvipalveluympäristössä useiden laitteiden vastuulle. MapReduce perustuu kahden operaation abstraktioon:

- *Map*: Prosessoi saamansa avain/arvo parit ja tuottaa näistä tuloksena nolla tai enemmän avain/arvo pareja.
- *Reduce*: Kutsutaan kerran jokaisen uniikin avaimen kohdalla. Funktio iteroi kaikki sellaiset arvot lävitse, jotka jakavat saman avaimen ja tuottaa tuloksena nollan tai nollaa isomman arvon.



Kuva 8. Simuloitu kuvaus MapReduce –funktion toiminnasta.

(<http://www.edureka.co/big-data-and-hadoop-course-curriculum>)

MapReducen suoritus alkaa Map-vaiheella, jossa jokainen avain/arvo pari luodaan annetun syötteen pohjalta. Tämän jälkeen *Shuffle*-vaiheessa edellisen vaiheen avain/arvo parit ryhmitellään avaimen mukaisesti (saman avaimen omaavat samaan ryhmään). Lopuksi Reduce-operaation avulla avain/arvo pareista koostetaan lopullinen tulos yhdistämällä arvot yhtenevän avaimen alaisuuteen. Tarkemmin operaation suoritus on kuvattu kuvassa 8, jossa simuloidaan sanojen määrän laskemista saadusta syötteestä.

Hadoopin ja MapReducen käytön avulla vältetään myös hajautetun ohjelmasuorituksen, datan hajauttamisen ja jakamisen sekä virheistä palautumisen ongelmiin varautumiselta, mitkä tarjotaan suoraan Hadoop-kirjaston kautta. Nämä yhdessä tekevät MapReducesta yksinkertaisen, mutta

tehokkaan vaihtoehdon suurten datajoukkojen käsittelyyn yhdistettynä järjestelmän riippumattomuuteen syötteenä annetun datan määrästä [Vo *et al.*, 2011]. Vo ja kumppanit myös huomioivat tutkimuksessaan, että yleisiä visualisointiin käytettyjä algoritmeja voidaan luonnollisesti kuvata MapReducen käyttämällä abstraktiolla ja luoda täten yksinkertaisia sekä erittäin hyvin skaalautuvia järjestelmiä.



## 4. VIRTUAALITODELLISUUDEN HYÖDYNTÄMINEN

### 4.1 Virtuaalitodellisuus aikaisemmin

Tieteellisessä tutkimuksessa virtuaalitodellisuus on jo pidempään nähty tehokkaana alustana ihmisen ja teknologian välisessä vuorovaikutuksessa [Sutherland, 1968; Cruz-Neira *et al.*, 1993; Burdea & Coiffet, 2003]. Kim [2005] määritteli tutkimuksessaan virtuaalitodellisuuden teknologiaksi, joka takaa saumattoman käyttökokemuksen ja koostuu useiden näyttöjen kokoonpanosta. Näiden tehokkaiden ja innovatiivisten laitteiden onkin todettu tarjoavan työkalu moniulotteisen ja kollaboratiivisen datan visualisoimiseen. Wangin ja kumppaneiden [2015] mukaan virtuaalitodellisuudella pystytään erityisesti edistämään tiedon geometristä ymmärrystä (muotojen ja kokojen merkitys) sekä hahmottamaan tiedon sisältö intuitiivisemmin tehokkaamman visualisoinnin kautta.

CAVE-ympäristöjä (*Cave Automatic Virtual Environment*) on jo pitkään hyödynnetty osana tieteellisiä tutkimuksia, joissa on haluttu hyödyntää virtuaalitodellisuuden elementtejä. CAVE-pohjaiset järjestelmät ovat olleet hyvin kalliita ja niiden pystyttäminen on vaatinut paljon aikaa. Esimerkiksi Oculus Rift ja HTC Vive tarjoavatkin nykyään CAVE-ympäristöjä halvemmän ratkaisun hyödyntää virtuaalitodellisuutta eri toimialueilla ja sen mahdollisuudet on huomattu myös tutkimuspuolella. Ensimmäinen HMD-tyyppinen (*Head Mounted Display*) laite kehitettiin Sutherlandin [1968] tutkimuksessa ja Cox, Patterson sekä Thiebaut [Cox *et al.*, 1997] lähettivät patentin liittyen äänen sekä eleiden hyödyntämiseen 3D-ympäristöissä. Cruz-Neira ja kumppanit [Cruz-Neira *et al.*, 1993] hyödynsivät jo aikaisessa vaiheessa CAVE-ympäristöä tutkimuksessaan ja Beck [2003] hyödynsi virtuaalitodellisuutta omassa kaupunkija mallintavassa VRGIS-järjestelmässä, joka tunnetaan tutkimusalalla hyvin. Myöhemmin Foo ja kumppanit [2009] ottivat virtuaalitodellisuuden mukaan myös terveydenhuoltoon ja käyttivät virtuaalitodellisuutta osana endoskooppisten operaatioiden suunnittelua. Näiden lisäksi virtuaalitodellisuuden tutkimus on vahvasti pyrkinyt löytämään ratkaisuja vaikeista motorisista vammoista kärsivien ihmisten kuntoutukseen. Hyödynnetyn teknologian näkökulmasta katsottuna CAVE-tyyliset ratkaisut ovat tällä hetkellä tutkimuksissa korvautumassa HMD-pohjaisilla virtuaalitodellisuuden laitteilla, mikä on varsin ymmärrettävää. Virtuaalitodellisuuden ala on kuitenkin hyvin uusi ja sen potentiaali on vielä nykyäänkin huonosti hyödynnetty. Tästä kertoo virtuaalitodellisuusalustoille tuotetun sisällön ja ratkaisujen niukkuus varsinkin pelikategorian ulkopuolella. Tutkimuksellisessa mielessä tilanne on tietysti hyvä, koska se avaa mahdollisuuden pohtia, mitä kaikkea virtuaalitodellisuudella voitaisiin saada aikaan. Alan kehityksen myötä virtuaalitodellisuuslasit mahdollistavat jatkossa myös

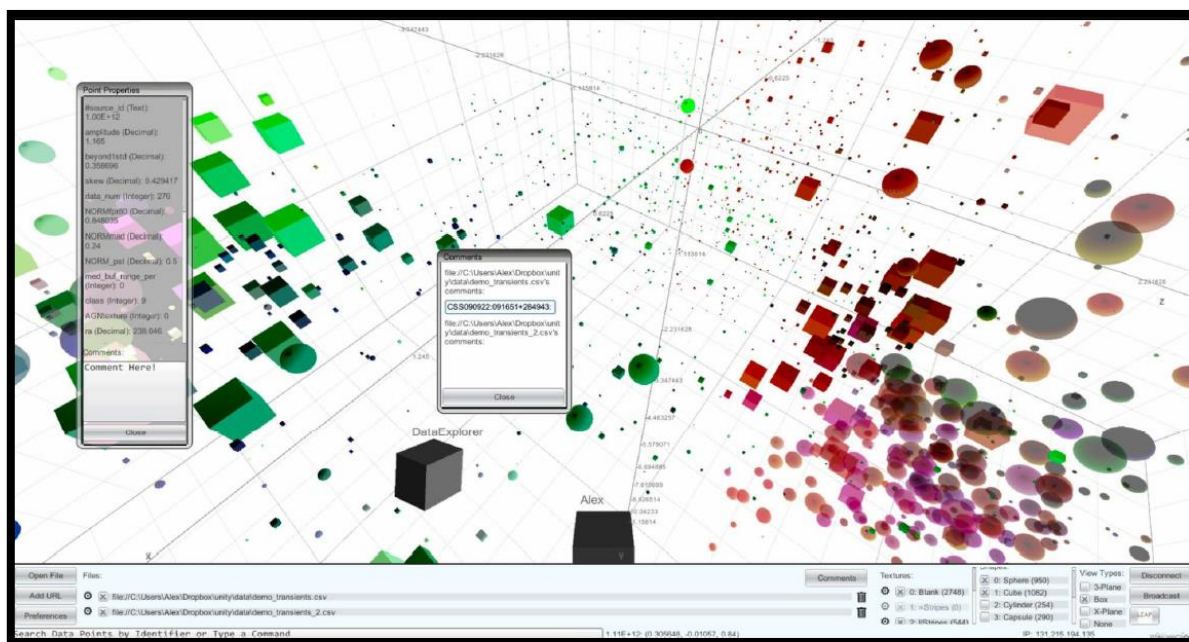
katseenseurannan, minkä hyödyntäminen on erityisen arvokasta. Tällä hetkellä virtuaalitodellisuuteen pohjautuvassa sisällöntuottamisessa on haasteena erityisesti käyttäjän vapaus katsoa mihin haluaa, jolloin kriittinen sisältö saattaa mennä ohitse. Tällöin katseenseurannan avulla voidaan tarjota käyttäjälle huomioita esitettävään sisältöön liittyen, jos käyttäjän huomio on kiinnittynyt toisaalle.

Virtuaalitodellisuuden esittelyn jälkeen tässä työssä käydään yleisesti lävitse virtuaalitodellisuuden HMD-pohjaisen teknologian sisältö ja teknologia sekä keskitytään erityisesti HTC:n keittämiin *Vive*-virtuaalitodellisuuslaseihin, joita tutkielmassa on tavoitteena hyödyntää.

## 4.2 Big Datan visualisointijärjestelmät

Soveltuva tiedon visualisointi on Tengin ja kumppaneiden [2015] mukaan suurin helpottava tekijä Big Datan hyödyntämisessä ja analysoinnissa. Myös Plugfelder ja Helmut [2013] pitävät kattavaa tiedonvisualisointia vaatimuksena sille, että kerätty tieto saadaan valjastettua käyttöön ja saatettua myös vähemmän asiantuntevuutta omaavien henkilöiden käyttöön. Virtuaalitodellisuutta hyödyntäviä Big Datan visualisointijärjestelmiä ei ole kaupallisina versioina saatavilla, mutta aiheeseen liittyviä tutkimuksia ja tutkimuksellisia järjestelmiä on tehty muutamia. Seuraavaksi käydään läpi kolme Big Datan visualisointiin pohjautuvaa virtuaalitodellisuuden järjestelmää ja arvioidaan niiden tutkimuksellista sisältöä tätä tutkielmaa silmällä pitäen.

Donalekin ja kumppanien [2014] tekemää tutkimusta voidaan pitää ensimmäisenä työnä, jossa virtuaalitodellisuuden hyötyjä Big Datan visualisoinnissa on tutkimuksellisesti arvioitu. He kehittivät tutkimuksessaan iVIZ-visualisointijärjestelmän, joka on suunnattu käytettäväksi tiedon analysoimista varten virtuaalilasien tai suoraan selaimen kautta. Donalekin ja kumppaneiden [2014] tutkimuksen lähtökohta oli hyvin samanlainen kuin tässäkin tutkimuksessa: Hyödyntää virtuaalitodellisuutta yleisenä, abstrahoituna visualisointityökaluna, joka tarjoaisi mahdollisuuden silmäillä ja analysoida mitä tahansa tietoa. Donalekilla ja kumppaneilla [2014] oli selvä näkökulma tutkimuksessaan siitä, että algoritmit eivät vielä nykyään pysty löytämään piilotettuja kytköksiä tietojoukon sisältä vaan tulee hyödyntää ihmisen luontaista havainnointikykyä. Tästä syystä iVIZ-järjestelmässä koko tietojoukko piirretään virtuaalimaailmaan ilman esisuodatusta. Järjestelmässä dataelementit kytketään osaksi XYZ -akselistoa, johon käyttäjä pystyy tekemään omia määrittäisiään. Käyttäjälle annetaan mahdollisuus tutkia sisältöä virtuaalitodellisuudessa täysin vapaasti eri suodatusmekanismeja hyväksikäyttäen.



Kuva 9. Donalekin ja kumppaneiden [2014] kehittämän iVIZ-järjestelmän perusnäkymä.

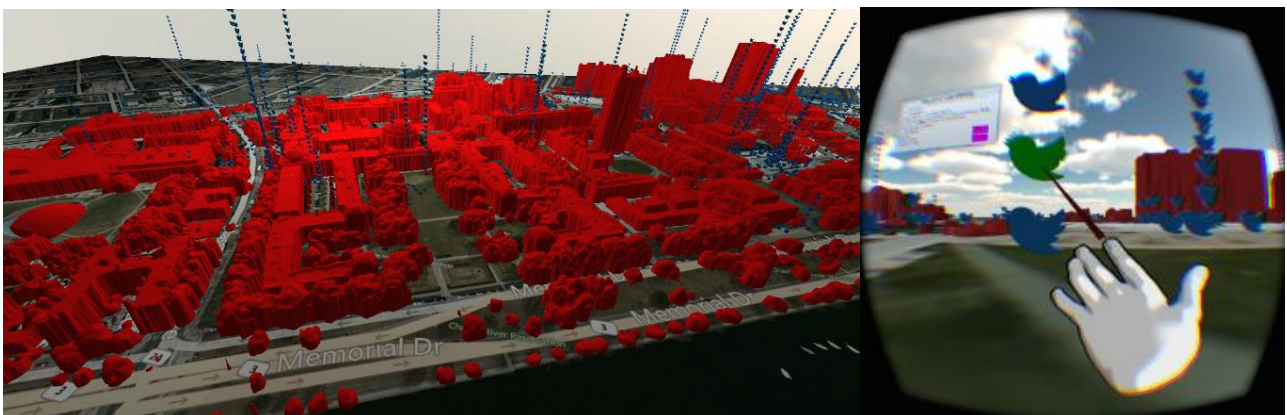
Tutkimuksessa löydettiin viisi eri näkökulmaa, miten tietoa voitaisiin esittää mahdollisimman kattavasti osana yhtä dataelementtiä:

- *XYZ* – Sijainti XYZ-koordinaatistossa
- *RGBA* – punainen, vihreä, sininen, alpha väritasokoodaus
- *Koko* – Elementin säteen pituus
- *Muoto* – Kuutio, pyramidi, kolmio, sylinteri, pallo
- *Tekstuuri* – Kuva, joka voidaan piirtää elementin pintaan

Vaikka iVIZ-järjestelmän arvioinnissa järjestelmän osoitettiin parantavan tiedon analysoinnin prosessia, järjestelmää ei silti voi kuvata kovin innovatiiviseksi. Donalek ja kumppanit [2014] eivät tutkimuksessaan avanneet tai perustelleet tekemiänsä valintoja järjestelmän visualisointipäätöksiin liittyen. Tutkimuksen pohjalta voidaan vain todeta, että iVIZ piirtää dataelementtejä näkyviin hyödyntäen pisteparvi-visualisointia (*Scatter plot*), joka erityisesti ison datajoukon kohdalla tekee visualisoinnista sekavan. iVIZ-järjestelmästä tarjolla olevat kuvankaappaukset itsessään jo osoittavat, että pisteparvi-tekniikka toimii heikosti myös virtuaalitodellisuutta hyödynnettäessä (Kuva 9). Donalek ja kumppanit mainitsevat tutkimuksensa lopuksi julkaisevansa alustan tiedeyhteisön käyttöön. Vaikuttaa kuitenkin siltä, että järjestelmän kehitystyö on lopetettu.

Toinen tieteellistä tutkimusta varten luotu Big Datan virtuaalitodellisuuden visualisointialusta luotiin Moranin ja kumppaneiden [2015] tutkimustyössä. Heidän tutkimuksessaan

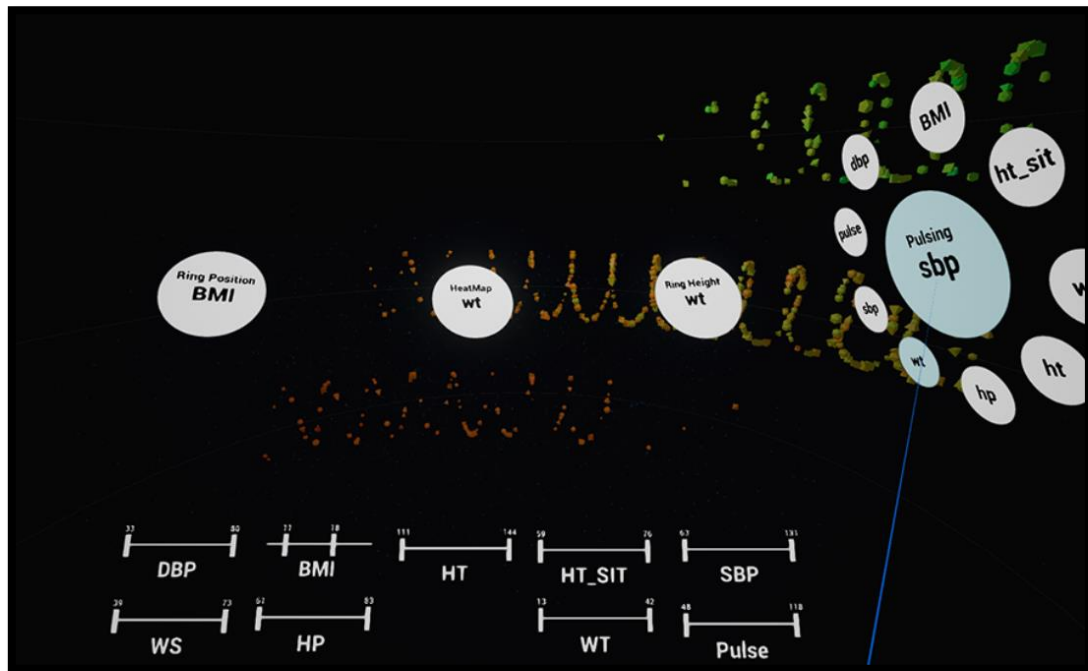
oli tavoitteena visualisoida tietoa MIT-kampuksen alueella lähetettyjen Twitter viestien sisällön pohjalta. Tutkimusryhmä mallinsi virtuaalitodellisuuteen kampusalueen ja Twitter-viestien geolokaatio-metatietoa hyödyntämällä he pystyivät sijoittamaan viestin lähetyspaikan osaksi 3D-maailman sijaintia. Heidän tutkimuksessaan visualisoinnille ei asetettu muita määritteitä kuin data-elementin sijainnin luodussa 3D-maailmassa pohjautuen mainittuun metatietoon. Täten visualisoinnin data elementit ovat vain kasattu päällekkäin niiden jakaessa saman geolokaation 3D-mallinnuksen sisällä (Kuva 10). Moranin ja kumppaneiden tutkimuksellisenä ongelmana on se, että he eivät varsinaisesti työstäneet Big Dataa tutkimuksessaan. Järjestelmässä kuvattiin vain Twitter-viestien sisältöä, joiden tietomalli on hyvin tunnettu, minkä myös heidän tutkimus vahvistaa: vain 2 % viesteistä sisälsi puutteita ja nämä jätettiin tutkimuksen ulkopuolelle. Tämän lisäksi tutkimuksessa käytetyn datajoukon määrä oli vain 6000 Twitter-viestiä ja data ei ollut reaaliaikaista. Näistä syistä myöskään heidän tutkimuksensa ei vastaa Big Datan visualisointiin liittyviin peruskysymyksiin eikä tutkimuksessa ole otettu kantaa näihin kysymyksiin miltään osin.



Kuva 10. Moranin ja kumppaneiden [2015] visualisointialustan näkymät.

Vuonna 2015 järjestetyn Big Data VR haasteen voittajaryhmä *Masters of Pie* kehitti oman Big Datan visualisointiin pohjautuvan järjestelmänsä (Kuva 11), joka osoittaa aikaisempia tutkimuksia paremmin visuaaliset ja vuorovaikutteiset mahdollisuudet, joita virtuaalitodellisuudessa voitaisiin hyödyntää. Ryhmä huomasi jo kehitysprosessin alussa, että pelkästään datan ripottelu virtuaalitodellisuuteen, kuten Donalekin ja kumppaneiden [2014] sekä Moranin ja kumppaneiden [2015] tutkimuksissa, ei ole merkityksellistä. Virtuaalitodellisuuteen luodun toteutuksen täytyisikin jo itsessään helpottaa analyysin tekemistä. Ryhmä päätyi visualisoinnissa DNA-ketjuun pohjautuvaan ratkaisuun, jossa tietosisältö kiertyy spiraalimaisesti käyttäjän ympärille ja tuo kaiken tiedon suoraan käyttäjän näkyville. Masters of Pien ratkaisussa käyttäjälle ei tarjota mahdollisuutta liikkua vapaasti visualisoinnissa vaan luotetaan visualisoinnin toimivuuteen itsessään sekä tehokkaisuuteen.

vuorovaikutuksellisiin työkaluihin. Perinteisten valintojen ja suodatusten lisäksi heidän ratkaisussa tarjotaan mahdollisuus datan eri ulottuvuuksien linkityksiin, jonka avulla käyttäjällä on mahdollisuus pyrkiä löytämään eri klustereita data joukon sisältä annettujen määritysten pohjalta. Masters of Pien tekemän toteutuksen taustatekijöitä esimerkiksi käytetyn datan osalta ei ole avattu, mutta heidän selvitystensä pohjalta käy selväksi, että käytettävän datajoukon sisältö on ollut etukäteen selvillä ja ainakin joiltain osin visualisointia on räätälöity kyseisen datajoukon mukaiseksi.



*Kuva 11. Masters of Pien visualisointityökalun toteutus, jossa kuvattuna on eri arvojen linkittäminen osaksi visualisoinnin eri ulottuvuuksia kuten kehän korkeus tai sijainti.*

#### 4.3 Yhteenveto aikaisempien järjestelmien pohjalta

Esimerkkien läpikäymisen jälkeen voidaan todeta, että Masters of Pie -ryhmän visualisointiratkaisu vaikuttaa tehokkaimmalta ja he ovat ratkaisussaan käyttäneet omaa innovaatiotaan. Järjestelmässä on keskitytty hyvään visualisointiin ja vuorovaikutuksellisiin työkaluihin, kun taas muiden alustojen ratkaisu pohjautuu vain datan sijoittamiseen maailmaan ja vapaaseen liikkuvuuteen sen ympärillä. Vapaan liikkumisen periaatetta ei voida nähdä datan analysoinnissa välttämättä perusteltuna, sillä analysoinnin vaiheessa, jota kyseiset järjestelmät palvelevat, halutaan nähdä datajoukon kokonaiskuva eikä niinkään olla kiinnostuneita yksittäisten elementtien arvoista. Täten erityisesti pääklustereiden löytäminen on datan käsittelyn alkuvaiheessa yksi tärkeimmistä tekijöistä, mikä voidaan ainakin osittain nähdä toteutuvan Masters of Pien toteutuksessa järjestelmän tarjoamien työkalujen avulla. Vapaa liikkuminen virtuaalitodellisuudessa voidaan myös nähdä mahdollisuutena



vaihtaa näkökulmaa tietojoukkoa tarkasteltaessa. Donalekin ja kumppaneiden [2014] sekä Moranin ja kumppaneiden [2015] tutkimuksissa data on kuitenkin sijoiteltuna maailmaan matriiseja hyödyntäen, jolloin kuvakulman vaihtamisen merkitystä on vaikea perustella tietoalkioiden ollessa staattisesti kiinnittyneinä koordinaatistoon.

Yhtenevää näillä kaikilla esimerkeillä on siinä, että yhdessäkään toteutuksessa ei oteta täysin kantaa tämän tutkimuksen alkuvaiheessa esitettyihin Big Datan hyödyntämisen ongelmakohtiin: Data ei ole reaaliaikaista, käytetyn datan sisältö on hyvin tai ainakin joiltain osin tiedossa ja datamäärät ovat hyvinkin maltillisia. Näiden tekijöiden takia aikaisemmat visualisointialustat tarjoavat heikon pohjan tässä tutkimuksessa tehtävää toteutusta ajatellen sekä yleisesti jatkotutkimusta silmällä pitäen.

Näiden esimerkkien pohjalta voidaan kuitenkin varmentaa jo aikaisemmin esille tullut tieto siitä, että Big Datalle suunnatun geneerisen ja datan sisältöön mukautuvan virtuaalitodellisuusjärjestelmän luominen on hyvin hankala prosessi. Näiden näkemysten pohjalta ei voida ajatella, että järjestelmä osaisi algoritmien pohjalta muodostaa aina oikean visualisoinnin annetulle datasyötteelle. Käyttäjän toiminnan ja tiedon välisen vuorovaikutuksen (suodatus, attribuuttien linkitykset) tulee olla pääpainotettuna myös visualisoinnin luomisessa, jos visualisoinnin prosessi halutaan mahdollistaa riippumatta datan sisällöstä. Tässä ajatuksessa palataan jälleen näkemykseen, että ihminen omaa erinomaisen kyvyn havaita datan sisäisiä rakenteita (*pattern*) ja riippuvuuksia näköaistinsa avulla, mikä koneellisesti vaatisi syvää prosessointia. Tällöin käyttäjälle annetaan vapaus poistaa häiriöt ja ottaa tarkasteluun vain häntä kiinnostavat osuudet, minkä turvin osa Big Datan visualisointia koskevista ongelmista saadaan eliminoitua. Edellä mainitut tekijät tullaan huomioimaan tämän tutkimuksen järjestelmän kehityksessä, joka käydään tarkemmin läpi luvussa 5. Lisäksi luvussa käydään läpi kompromissit sekä niiden perustelut, mitä mahdollisesti joudutaan kehityksen aikana tekemään eri riippuvuustekijöistä johtuen.

#### 4.4 HTC Vive

Vive on HTC:n kehittämä virtuaalitodellisuuslasien teknologia. Järjestelmä tarjoaa tällä hetkellä kokonaisvaltaisemman virtuaalitodellisuusratkaisun kuin esimerkiksi kilpailija Oculus Rift. HTC Vive tarjoaa virtuaalitodellisuusnäkyvän lisäksi käyttäjälle mahdollisuuden siirtää oma liikehdintä reaali maailmassa suoraksi liikkeeksi virtuaali maailmassa jäljentämällä käyttäjän liikkumista erikseen määritellyn alueen sisällä infrapunakameroiden avulla. Lisäksi järjestelmä pystyy seuraamaan päässä olevien lasien ja käsissä olevien ohjainten liikehdintää. Taulukkoon 2 on koottu vertailu markkinoilla olevien virtuaalilasien ominaisuuksista. Taulukon pohjalta voidaan todeta, että HTC Vive on

vaihtoehtoista tällä hetkellä monipuolisin ja tarjoaa suoraan ratkaisun interaktioon virtuaalimaailman sisällä ilman ulkopuolisten laitteiden kytkemistä osaksi järjestelmää.

Virtuaalilasit 21.3.2016 Virtuaalimaailma.fi	Google Cardboard	Samsung Gear VR	HTC Vive	Oculus Rift	Sony PSVR
Hinta	7€	n. 150€	n. 950€	n. 750€	n. 500€
Langaton	kyllä	kyllä	ei	ei	ei
Ohjaimet	ei	Kehitteillä	Käsiohjaimet mukana	Xbox ohjain, käsiohjaimet Q2 2016	Käsiohjaimet mukana
Liikkuminen VR-tilassa	ei	ei	5 x 5 m	1,5 x 1,5 m	1,5 x 1,5 m
Resoluutio	Riippuu puhelimesta	1280x1440	1080x1200	1080x1200	1080x960
Kuvataajuus	Riippuu puhelimesta	60	90	90	120
Peligrafiikan laatu	*	**	****	****	****
Parhaimmillaan	360 video	360 video	Pelit huoneen kokoisessa tilassa	Pelit istuen tai seisoen	Pelit istuen tai seisoen
Julkaisu	Myynnissä	Myynnissä	05/04/2016	28/03/2016	H1 2016
Vaatii toimiakseen	Älypuhelimien	Uudehkon Samsung puhelimen	Tehokkaan tietokoneen	Tehokkaan Tietokoneen	PS 4 pelikonsolin

*Taulukko 2. Markkinoilla olevien virtuaalilasien ominaisuuksien vertailu.*

*([www.virtuaalimaailma.fi/virtuaalilasit/](http://www.virtuaalimaailma.fi/virtuaalilasit/))*

Aikaisemmissa, virtuaalitodellisuuteen pohjautuvissa visualisointijärjestelmissä, on hyödynnetty Oculus Rift -virtuaalilasiteknologiaa. Oculus Rift -virtuaalilasien lisäksi Donalekin ja kumppaneiden [2014] tutkimuksessa hyödynnettiin *Vicon* liikkeenseurantajärjestelmää [Vicon, 2018], jonka avulla käyttäjän liike siirrettiin osaksi virtuaalista liikettä. Valintojen ja syötteiden antamiseen kaikissa tutkimuksissa [Donalek *et al.*, 2014; Moran *et al.*, 2015; Masters of Pie, 2015] hyödynnettiin *Leap Motionin* liikkeiden ja eleiden tunnistusjärjestelmää [Leap Motion, 2018]. Kuten mainittu, nämä vuorovaikutukselliset elementit ja teknologiat huomioiden Vive tarjoaa suoraan vaaditut ominaisuudet. Lisäksi Viven ominaisuudet voidaan helposti ottaa käyttöön Unity3D-kehitysympäristöön suunnattujen kirjastojen avulla, jolloin alustan kehitysaikaa säästyy. Näistä syistä tämän tutkimuksen alusta tullaan lähtökohtaisesti kehittämään Viven teknologian pohjalle. HTC Viven käyttöä tukee myös Masters of Pie -ryhmän artikkelissa tekemä toteamus, jossa he toteavat, että alustan käyttökokemus olisi parempi HTC Viven tekniikalla toteutettuna kuin Oculuksen laitteistolla.

## 5. TESTIJÄRJESTELMÄN KEHITYS

Testijärjestelmän kehittämisen tarkoituksena oli käyttäjätestauksessa kerätä testikäyttäjiltä tuntemuksia ja mielipiteitä tiedon visualisoimisesta virtuaalitodellisuudessa ja siitä, miten visualisointitapa vertautuu tietokoneen näytöltä esitettyyn vastaavaan visualisointiin. Kehityksessä tavoitteena oli luoda datan visualisoinnin toteutus virtuaalitodellisuusympäristöön sekä vastaava 2D-toteutus työasemaympäristöön.

Visualisointidemoja luotiin kolme erilaista: Interaktiivinen visualisointi maapallosta, johon reaaliaikaisesti päivittyy näkyville Twitter-viestejä näkyville niiden geolokaatitietoihin pohjautuen. Lisäksi toteutettiin kaksi perinteisempää visualisoinnin toteutusta pylväsdiagrammista ja graafista. Jokaisella visualisointitavalla on tutkielmassa oma roolinsa, minkä takia kyseinen visualisointitapa on käyttäjäutkimukseen valittu:

- Maapallo-demo edustaa räätälöidympää ja dynaamisempaa visualisointia, jossa data on sidottu vahvasti näkymän kontekstiin. Visualisoinnin pitäisi luoduista demoista pystyä tarjoamaan vahvimman immersion.
- Pylväsdiagrammi edustaa perinteistä ja työasemaympäristön käytössä totuttua visualisointitapaa, jossa palkit on diagrammissa aseteltu toisiinsa nähden lomittain, jolloin 3D-maailman syvyysaspekti tulee käytössä ottaa huomioon.
- Graafi-demossa on poistettu käyttäjän mahdollisuus vuorovaikuttaa visualisoinnin kanssa, jolloin demo keskittyy kaikista yksinkertaisimmillaan vertaamaan kahden eri visualisointinäkymän kokemuksellisuutta ilman, että ympäristöjen välillä eroavat kontrollit vaikuttavat kokemukseen.

Testijärjestelmän kokonaisuudet kehitettiin Unityn 3D-pelimoottorilla, joka erityisesti tukee järjestelmien kehitystä useille eri alustoille sekä tekee virtuaalitodellisuuden mallintamisen helpoksi [Unity, 2018]. Unityn avulla työasemaympäristön demojen visualisointi on siirretty täysin virtuaalitodellisuutta vastaavana näkymänä, mutta järjestelmän ja käyttäjän välisessä vuorovaikutuksessa on alustasta johtuvia eroavaisuuksia. Nämä toiminnollisuuden eroavaisuudet on kuvattu demokohtaisesti kappaleissa 5.1.1 ja 5.2.1. Kappaleessa 4.4 arvioitiin HTC Vive -virtuaalitodellisuuslasien olevan tällä hetkellä paras vaihtoehto virtuaalitodellisuusympäristön käyttämiseen. Tutkimusta varten Viven laseja ei kuitenkaan saatu käyttöön. Tästä syystä testijärjestelmät jouduttiin luomaan saatavilla olleille Oculus Rift DK2 -laseille. Oculus-sarjan ohjaimia ei myöskään saatu tutkimuksessa käyttöön, joten virtuaalitodellisuuden demoissa jouduttiin järjestelmän ohjaamisessa hyödyntämään Playstation 4 -ohjainta. Käyttäjätestauksia varten ei



myöskään päästy hyödyntämään tehotyöasemaa vaan testit jouduttiin suorittamaan Lenovo ThinkPad W541 kannettavalla tietokoneella. Tästä syystä saatavilla olevan laitteiston tehojen puutteella on vaikutusta testauksessa läpikäytävien demojen suorituskykyyn erityisesti virtuaalitodellisuuden näkymien osalta. Lisäksi hyödynnettävissä olevien laitteistokompromissien takia virtuaalitodellisuuden demoihin ei saatu luotua sitä käyttäjäkokemusta, joka olisi ennakoon haluttu. Näiden tekijöiden vaikutukset käyttäjätutkimuksessa näkyivät ja niiden merkitystä on tarkemmin käyty lävitse käyttäjätestin tuloksissa kappaleessa 6.2.

Kahden luodun testijärjestelmän sisältö sekä toteutusvaihe käydään seuraavaksi tarkemmin lävitse. Kummastakin testijärjestelmästä annetaan tarkempi yleisselvitys, käydään käyttäjätestauksessa hyödynnettävät toiminnollisuudet lävitse sekä peilataan tehtyä toteutusta teoriaosuuden luvussa 2 Big Datalle annettuja määritteitä vasten. Osiossa arvioidaan kuinka hyvin järjestelmä vastaa Big Datan annettuja määritteitä ja mitä näistä määritteistä ei ole pystytty tässä työssä lunastamaan. Määritteiden perusteluissa myös selvitetään, minkälaisilla muutoksilla kyseisen osion saisi vastaamaan Big Datan viiden V:n (Volume, Velocity, Variety, Veracity, Value) vaatimuksia.

### **5.1 Twitter-viestien visualisointi maailmankartalla**

Yhden demojärjestelmän sisällöksi haluttiin luoda staattisemmasta visualisoinnista poikkeava interaktiivisempi toteutustapa. Demon (Kuva 12) tarkoituksena on luoda syvempi visualisoinnillinen kokemus, jossa pystytään paremmin hyödyntämään virtuaalitodellisuuden interaktiotapoja kuten pään liikkeenseuranta. Visualisoinnissa haluttiin hyödyntää laajaa tietomallia, joka päivittyisi dynaamisesti käytön aikana. Lisäksi tietosisällön tuli olla monimuotoista, jotta visualisoinnilla pystyttäisiin tukemaan datan sisällön monimuotoisempaa esittämistapaa. Näiden kriteerien pohjalta visualisoitavaksi dataksi valikoitu Twitter-viestit, joita työssä hyödynnetään *Twitter Streaming API* rajapinnan avulla [Twitter Stream API, 2018]. Twitter-viestien rajapinta tarjoaa mahdollisuuden suoratoistaa ja kuunnella Twitteriin luotuja viestejä reaaliajassa (noin 6000 viestiä sekunnissa). Rajapinta tarjoaa täyden viesteihin liittyvän tietosisällön esimerkiksi viestin sisällön ja käyttäjänimen.

Käytettävään dataan ja tiedonhankintaan liittyvien päätösten jälkeen seuraava vaihe oli miettiä, minkälainen visualisointi datan ympärille tulisi luoda. Twitter-viestien visualisointi viestien geolokaatitietojen pohjalta maailmankartalle sijoiteltuna on tehty jo aikaisemmin #Onemilliontweetmap-projektissa [Onemilliontweetmap, 2018] ja vastaavan rajapinnan hyödyntämistä sivuttiin kappaleessa 4.2 esitetyn Moranin ja kumppaneiden [Moran *et al.*, 2015]

visualisointijärjestelmässä, mutta kyseinen rajapinta tarjosi hyvän lähtökohdan yksityiskohtaisemman visualisoinnin luomiselle. Viestien visualisointi maapallon päälle tuntui myös tarjoavan syvällisemmän käyttökokemuksen virtuaalitodellisuuteen. Maapallon 3D-malli ja sen toiminnollisuudet otettiin käyttöön Unitylle luodusta World Political Map -kirjastosta [World Political Map, 2018]. Tämän jälkeen Twitter viestien käsittelyä varten luotiin oma Golangilla toteutettu API-rajapinta (*Application Programming Interface*), joka säilöö halutut viestitiedot tietokantaan. Viimeisessä vaiheessa rajapintaan toteutettiin tarvittavat tietokyselyt Unity-projektin haettavaksi. Näiden tietojen pohjalta elementit pystytään Unityssa mallintamaan geolokaatiotietojen pohjalta oikeille paikoille ja päivittämään maapallolla esitettävät viestit muutaman sekunnin välein.



*Kuva 12. Havainnekuva Twitter-viestien maapallo visualisointidemosta.*

### **5.1.1 Toiminnollisuus**

Visualisoinnissa tärkeimpänä tietona korostuvat Twitter-viestin sijaintikoordinaatit. Koordinaattien avulla viestit piirretään niiden lähettämispaikkaa vastaavaan sijaintiin maapallon 3D-mallissa. Käyttäjä voi tarkistella viestien sisältötietoja viemällä ohjattava kursori viesti-ikonin päälle. Kursorin ohjaaminen tapahtuu virtuaalitodellisuuden toteutuksessa virtuaalitodellisuuslasien liikeseurannan

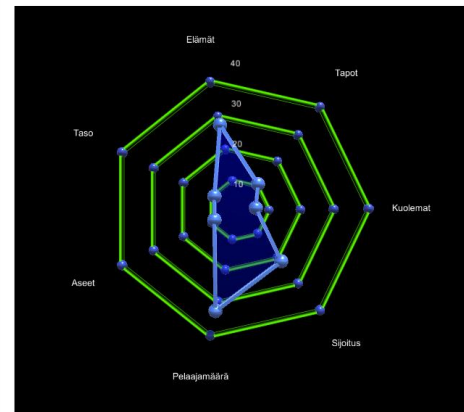
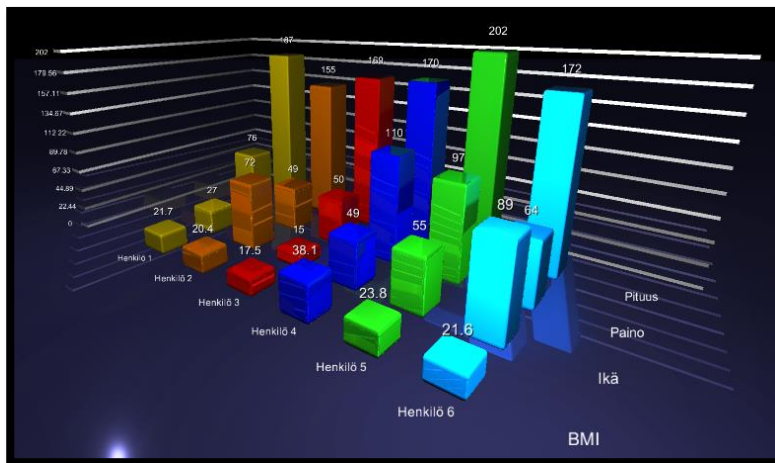
kautta ja työpöytätoeuteuksessa hiiren avulla. Käyttäjä pystyy myös kääntämään maapalloa virtuaalitodellisuuden toeuteuksessa erillisen ohjaimen avulla sekä työpöytäympäristössä hiiren painikkeen avulla. Virtuaalitodellisuuden ja työasemaympäristöjen vuorovaikutukselliset eroavaisuudet on kuvattu taulukossa 3.

Interaktio	Virtuaalitodellisuus	Työpöytäympäristö
<b>Kursorin liikuttaminen</b>	Toteutettu virtuaalitodellisuuslasien liikkeenseurannan avulla. Kursori liikkuu maapallon pinnalla päätä liikuteltaessa.	Kursorin liikuttaminen hiiren avulla.
<b>Maapallon kääntäminen</b>	Toteutettu PS4-ohjaimen avulla. Ohjaimen vasemman analogisen sauvan avulla maapalloa mahdollista pyörittää haluamaansa suuntaan.	Pitämällä hiiren painiketta pohjassa ja vetämällä maapalloa haluttuun suuntaan.
<b>Maapallon pyörittäminen paikallaan</b>	Ohjaimen X-painikkeen avulla mahdollista pyörittää mallia paikallaan, jotta maapallon saa helpommin käännettyä haluttuun kulmaan	Hiiren oikean painikkeen avulla mahdollista pyörittää mallia paikallaan, jotta maapallon saa helpommin käännettyä haluttuun kulmaan

*Taulukko 3. Virtuaalitodellisuuden ja työpöytäympäristön eroavaisuudet maapallovisualisoinnin demossa.*

## 5.2 Twitter-viestitietojen visualisointi pylväsdiagrammeilla ja graafilla

Toeuteuksen toisessa osassa haluttiin keskittyä perinteisempien visualisointitapojen käsittelyyn. Erityisesti kiinnostus testijärjestelmässä kohdistuu siihen, miten käyttäjät reagoivat kahteen käytännössä samanlaiseen diagrammiin, mutta virtuaalitodellisuuden ympäristössä nämä ovat tuotu ympäristön takia lähemmäksi käyttäjää. Esimerkkisovellusta varten päämääränä oli myös luoda kaksi erilaista diagrammimallinnusta, jotka eivät olisi toeuteukseltaan datan kontekstiin sidottuja, kuten maapallovisualisoinnissa, jossa tietosisällön koordinaattitiedot johdettiin visualisoinnissa suoraan maapallon paikkatietoon.



Kuva 13. Pylväsdiagrammin visualisointidemosta sekä Kuva 14. graafin visualisoinnista.

Taustateorian pohjalta kappaleessa 3.2.2 Big Datan visualisoimiseen esitettiin käytettäväksi tähtikoordinaattien visualisointia, joka soveltui erityisesti käytettäväksi dataan tutustumisessa data-analyysin ensimmäisessä vaiheessa ja visualisointi mukautui myös hyvin laajoihin datamääriin. Tässä työssä vastaan tulleista aikarajoitteista ja tähtikoordinaattien haastavammasta toteutustavasta johtuen visualisointia ei kuitenkaan pystytty 3D-mallintamaan virtuaaliodellisuuteen, joten sitä ei taustateoriasta huolimatta voitu tässä työssä hyödyntää. Diagrammeja varten työssä otettiin käyttöön Unityn kaupasta löytynyt kirjasto Graph And Chart [Graph And Chart, 2018], joka tarjoaa valmiina 3D-mallinnettuja diagrammeja hyödynnettäväksi virtuaaliodellisuudessa. Näistä saatavilla olleista malleista tehtiin esimerkkitoiteutuksia käyttäjätuesta varten pylväsdiagrammista (Kuva 13) sekä graafista (Kuva 14).

### 5.2.1 Toiminnollisuus

Pylväsdiagrammi ja graafi visualisoinneissa käyttäjälle suunnattu toiminnollisuus liittyy maapallovisualisoinnin tavoin mallien, graafin sekä pylväsdiagrammien, pyörittämiseen sekä valintakursorin liikuttamiseen. Kohdistamalla valintakursori katseella graafin tai pylväsdiagrammin eri osien päälle käyttäjälle annetaan palautteena kyseisen kohteen sisältämä arvo ja arvon selvennys esimerkiksi "Paino: 86 kg". Maapallovisualisoinnin mukaisesti käyttäjä pystyy silmäilemään diagrammivisualisointia haluamastaan kulmasta ohjaimen kääntöpainikkeen avulla.

Interaktio	Virtuaalitodellisuus	Työpöytäympäristö
<b>Kursorin liikuttaminen</b>	Toteutettu virtuaalitodellisuuslasien liikkeenseurannan avulla. Kursori liikkuu mallien päällä päätä liikutettaessa.	Kursorin liikuttaminen hiiren avulla.
<b>Mallin pyörittämien</b>	Toteutettu PS4-ohjaimen avulla. Ohjaimen vasemman analogisen sauvan avulla mallia mahdollista kääntää haluamaansa kulmaan.	Pitämällä hiiren painiketta pohjassa ja kääntämällä mallia/diagrammia haluttuun suuntaan.

Taulukko 4. Virtuaalitodellisuuden ja työpöytäympäristön eroavaisuudet diagrammien ja graafien demossa.

### 5.3 Big Datan hyödyntäminen ja sen ongelmat testijärjestelmissä

Seuraavaksi käydään lävitse Big Datan hyödyntämiseen liittyvät tekijät ja haasteet testijärjestelmissä, kappaleessa 2.1 esitettyjen Big Datan viiden määritteen pohjalta. Kappaleessa 2.2 esitetyn Big Datan kuudennen määritteen, visualisoinnin, tekijöitä ei käydä tässä yhteydessä lävitse sen ollessa tutkimuksen varsinaisena kohteena käyttäjätutkimuksessa ja käyttäjätutkimuksen analysoinnissa. Määritteiden toteutuman tarkemmat kuvaukset löytyvät maapallovisualisoinnin osalta taulukosta 5 ja diagrammivisualisoinnin osalta taulukosta 6.

Attribuutti	Määrittäminen täyttyy	Kuvaus
Volume	Ei	Käytetty datamäärä ei ole niin laaja, että se ei mahtuisi käytettävän laitteen lokaaliin muistiin. Käytettävien viestien tietueet ovat kohtuullisen pieniä, joten ennen laitteen tallennuskapasiteettiä rajoitteeksi muodostuisi mallinnettavien elementtien määrä pelimootorissa.
Velocity	Kyllä	Tieto päivittyy visualisointiin reaaliaikaisesti. Uudet viestit päivitetään näkymään kahden sekunnin välein.
Variety	Ei	Käytetyn tiedon rakenne on hyvin tunnettu Twitterin rajapintaa hyödynnettäessä. Tästä syystä tiedon rakenne ei vaadi erillistä analyysiä tai metatason selvitystä.

Veracity	Ei	Tieto on käyttäjien itse luomaa tai käyttäjän tiedoista koneellisesti johdettua, jolloin kaikki tieto perustuu johonkin todelliseen. Tällöin tiedon epäluotettavuus ei muodostu ongelmaksi.
Value	Kyllä	Tuhansien reaaliaikaisesti päivittyvien viestien kohdalla pelkkä viestien sisältämä tietomassa on itsessään lähes hyödytön, läpikäymisen muuttuessa erittäin raskaaksi. Esimerkissä tämä tieto on valjastettu päätöksentekoprosessia tehostavaan muotoon visualisoinnin avulla, jossa eri viestisisällön omaavat viestit voidaan paikallistaa.

*Taulukko 5. Maapallovisualisaation kuvaus Big Datan määritteiden osalta.*

Attribuutti	Määrittäminen täyttyy	Kuvaus
Volume	Ei	Käytetty datamäärä ei ole niin laaja, että se ei mahtuisi käytettävän laitteen lokaaliin muistiin. Visualisointien osalta datamäärä voitaisiin kasvattaa lokaalin muistin ylittäviin määriin ja tallettaa palvelinympäristöön, jos yli teratavun kokoista datajoukkoa hyödynnettäisiin.
Velocity	Kyllä	Visualisoinneissa käytettävä tieto on reaaliaikaisesti päivittyvä datajoukon sisällön muuttuessa.
Variety	Ei	Käytetyn tiedon rakenne on tunnettu käytettyjen datajoukkojen kohdalla. Tuntemattoman sisällön omaavaa datajoukkoa ei voida diagrammi- ja graafipohjaisten visualisointien kohdalla hyödyntää, sillä tuntemattomat tietueet korruptoisivat ja väärentäisivät visualisoinnin sisällön.
Veracity	Ei	Visualisoinneissa käytetty data on noudettu luotettavasta tietokannasta, jolloin tiedon luotettavuuteen liittyvää ongelmaa ei ole.
Value	Kyllä	Käytetty datajoukko on jokseenkin luettavaa suoraan tai taulukkopohjaisessa esitysmuodossa. Visualisointien avulla pystytään tuomaan tiedon sisältämä arvo paremmin esille valitsemalla kohteet, joista erityisesti ollaan kiinnostuneita (pylväsdiagrammi) tai samanaikaisesti visualisoimaan kattavammin moniulotteista datajoukkoa (graafi).

*Taulukko 6. Pylväsdiagrammi- ja graafivisualisaation kuvaus Big Datan määritteiden osalta.*

Tutkimuksen testijärjestelmissä voitiin ottaa huomioon joitain Big Datan määrittelyyn liittyviä tekijöitä, mutta kuten myös kappaleessa 4.2 läpikäydyissä, aikaisemmissa Big Dataa ja virtuaalitodellisuutta käsittelevissä tutkimuksissa, näiden tekijöiden huomioon ottaminen osoittautui hyvin vaikeaksi. Kummankin testijärjestelmän osalta määritteet täytyivät kohtien Velocity ja Value osalta. Testijärjestelmissä reaaliaikaisesti päivittyvän datan hyödyntäminen ja visualisointiin siirtäminen toteutettiin työssä suoraviivaisesti Twitter Streaming API:n ja koneellisesti generoitavan tiedon avulla. Rajapintoja hyödynnettäessä niistä saatava data ja sen rakenne on aina hyvin selkeästi kuvattu, jolloin käytettävän datan sisällön tarkka analysoiminen ei muodostu ongelmaksi ja se jätetään tässä tutkimuksessa huomiotta, mikä tapahtui myös aikaisemmassa Moranin ja kumppaneiden tutkimuksessa [Moran *et al.*, 2015]. Tiedon visualisointijärjestelmiä luotaessa myös tiedon arvon esille tuominen voidaan todeta työssä täyttyvän, vaikkakin jo tietoa hyödynnettäessä on ollut tiedossa sen sisältö ja sisällön tärkeät elementit.

Tutkimuksen testausjärjestelmissä kolme Big Datan viidestä määritteestä jää kuitenkin niille annetun määrittelyn pohjalta toteutumatta. Tiedon määrä (Volume) luo kaksi erityistä ongelmaa alustalle, jonka tulisi pystyä visualisoimaan Big Dataa virtuaalitodellisuudessa. Perusvaatimus datan niin isosta määrästä, että datajoukko ei mahtuisi paikallisen koneen kovalevylle (ainakin yli teratavu), vaatisi tässä tapauksessa oman pilvipalvelualustan, johon hyödynnettävä data säilöittäisiin ja data noudettaisiin erillisten rajapintakyselyiden avulla. Tällaisen alustan hankkiminen tai luominen tutkielman piirissä ei ollut mahdollista. Lisäksi näin suuren tietomäärän mallintaminen Twitter-viesteinä maapallovisualisaatiossa olisi vaatinut huomattavat määrät enemmän laskentatehoa laitteistolta, joka nyt oli saatavilla. Lisäksi toteutuksessa käytetyn Unity-pelimoottorin resurssit olisivat mahdollisesti tulleet vastaan, kun aletaan puhumaan usean miljoonan viestin mallintamisesta ympäristöön. Valittujen tietolähteiden alemmasta datan määrästä ja skaalautuvan pilvialustan käyttämättömyydestä johtuen, testijärjestelmien kehityksessä ei ollut tarve hyödyntää kappaleessa 3.5 esiteltyä MapReduce -arkkitehtuuria isojen tietomäärien hallitsemiseen.

Hyödynnettävän tiedon tuntematon rakenne (Variety) ja tiedon epäluotettavuus (Veracity) nähtiin kehitystyön aikana kaikista suurimpina ongelmakohtina Big Dataan pohjautuvia järjestelmiä luotaessa. Rakenteeltaan tuntemattoman datan visualisoiminen tehokkaasti virtuaalitodellisuuden ympäristössä ei tämän taustatutkimus- ja kehitystyön pohjalta vaikuta merkityksellistä tai sen ratkaiseminen vaatii huomattavasti laajemman selvityksen kuin tässä työssä. Tutkimustyöllisestä näkökulmasta tuntemattoman tai epäluotettavan tiedon käsittelyssä ongelmaksi muodostuu määritteisiin soveltuvan data joukon löytäminen tai luominen. Näin laajan, rakenteeltaan tuntemattoman, sekä silti jonkin säännön mukaan tietoa generoivan data joukon löytäminen ei täten



tämän tutkimuksen osalta toteutunut. Testijärjestelmiin valitun datan kohdalla nämä ongelmat eivät myöskään realisoituneet, sillä rajapintoja, kuten Twitter Streaming API, hyödynnettäessä rajapinnasta saatava data ja sen rakenne on aina hyvin selkeästi kuvattu. Lähimmäksi tuntematonta tai epäluotettavaa data joukkoa varmasti päästäisiin laajojen järjestelmien käyttölokeissa ja niihin liittyvässä analytiikassa. Käyttölokien analytiikkaa onkin käsitelty useissa aikaisemmissa tutkimuksissa [Miranskyy *et al.*, 2016; Wu, 2017] ja käyttölokien analytiikan tarkempi läpikäyminen ei ole osa tämän tutkimuksen sisältöä.

Lopulta tutkielman määrittämien Big Datan ominaisuuksien huomioiminen tutkimuksen toteutuksessa ei onnistunut sen paremmin kuin kappaleessa 4.2 esiteltyjen aikaisempien visualisointijärjestelmien kohdalla.



## 6. TESTIJÄRJESTELMIEN KÄYTTÄJÄTESTAUS

Luvussa esitellään testijärjestelmien käyttäjätestauksessa hyödynnettävä materiaali, tehtävät ja haastattelukysymykset. Luvussa käydään myös läpi käyttäjätestauksen koko prosessi ja sen eri vaiheet. Käyttäjätestauksessa hyödynnetyt lomakkeet, tehtävät ja haastattelukysymykset löytyvät tarkemmin tutkimuksen lopussa löytyvistä liitteistä.

### 6.1 Testisuunnitelma

Käyttäjätestauksessa tarkoituksena on selvittää minkälaisena käyttäjät kokevat visualisointijärjestelmän käyttökokemuksen virtuaalitodellisuuden ja työasemaympäristön välillä. Käyttäjätestauksen osallistujamäärä on 10 henkilöä. Osallistujiksi pyritään valitsemaan sekoitus sellaisia henkilöitä, jotka eivät ole aikaisemmin kokeneet virtuaalitodellisuuden ympäristöjä missään muodossa sekä osallistujia, joilta löytyy ainakin yksi aikaisempi kokemus virtuaalitodellisuudesta. Muita osallistujiin kohdistettuja kriteereitä ei tutkimuksessa ole asetettu.

Käyttäjätestauksen aikana käydään läpi virtuaalitodellisuuteen sekä työasemaympäristöön luodut demot. Demojen suoritusjärjestyksessä ensimmäisenä on maapallovisualisointi, jota seuraa diagrammeihin pohjautuva demo. Valitun ympäristön suoritusjärjestys arvotaan ennen käyttäjätestin alkamista, jotta testin aikaisella oppimistekijällä ei ole vaikutusta ympäristöjen suoriutumisen vertailussa. Järjestyksen satunnaistamista varten hyödynnetään *Latin square* -algoritmia, joka takaa sen, että tutkimuksessa hyödynnettyjen eri järjestysten määrä on yhtenevä, mutta satunnainen [Latin Square, 2018]. Osioiden suorittamisen aikana käyttäjälle esitetään tehtäviä ja kysymyksiä sekä osion suorittamisen jälkeen häntä pyydetään vastaamaan muutamiin kysymyksiin.

Käyttäjätestin koko prosessi ja sen eri vaiheet on tarkemmin esitetty kaaviossa 1, jossa esitettyjen kysymys- ja tehtäväjoukkojen järjestys on yhä tarkemmin kuvattu tutkimuksen mahdollista toistamista varten. Lisäksi testin aikana esitettävät haastattelukysymykset ja tehtävät käydään seuraavaksi tarkemmin lävitse.



*Kaavio 1. Esimerkki käyttäjätetausprosessin kulusta ja vaiheista. Maapallodemoon ensimmäiseksi ympäristöksi on arvottu desktop ja tämän tehtäväjoukoksi 2. Pylväs- ja graafidemon ensimmäiseksi ympäristöksi on arvottu virtuaalitodellisuus ja tehtäväjoukko 1.*

### 6.1.1 Haastattelukysymykset

Jokaisen demo osion sisältämien tehtävien lisäksi osioiden suorittamisen jälkeen käyttäjälle esitetään muutamia kysymyksiä käyttökokemuksen selvittämiseksi. Kysymyksistä viisi on kvalitatiivisia, joissa käyttäjää pyydetään vapaasti kuvailemaan käyttökokemusta. Haastattelukysymyksissä haluttiin pitää painotus kuvailevissa kysymyksissä, jotta testikäyttäjien kokemukselliset mielipiteet tulisivat vahvemmin esille ja kysymykset eivät liikaa ohjaisi ajatusprosessia. Täten kvantitatiivisia kysymyksiä esitetään vain kaksi jokaisen demon sekä jokaisen demokokonaisuuden jälkeen. Kvantitatiivisten kysymysten avulla tiedustellaan jokaisen demo-osion käytön mielekkyyttä sekä käyttökokemuksen luontevuutta. Kysymysten vastaukset esitetään 10-portaisella Likert-asteikolla arvoilla 1-10, minkä avulla kyselyyn vastaaja voi esittää positiivisen tai negatiivisen mielipiteensä [Likert, 1932]. Kvantitatiivisten kysymysten tuloksia käytetään tutkimuksen lopussa ensisijaisesti virtuaalitodellisuuden ja työasemaympäristön kokemuksen arvioinnissa mahdollistaen myös osioiden välisen vertailun.

Kysymysjoukko 1.

- *Miltä osion käyttäminen tuntui? (vapaa kuvaus)*
- *Kuinka luonnolliselta järjestelmän käyttö tuntui? (asteikko 1 – 10)*
- *Kuinka miellyttävältä järjestelmän käyttö tuntui? (asteikko 1 – 10)*

Kysymysjoukko 2.

- *Kuinka tehokkaalta järjestelmän käyttö virtuaalitodellisuudessa tuntui? (asteikko 1-10)*

- *Kuinka tehokkaalta järjestelmän käyttö tietokoneen näytöllä tuntui? (asteikko 1-10)*

### Kysymysjoukko 3.

- *Mikä näistä kolmesta demosta: Maapallo, pylväsdiagrammi tai graafi toimi mielestäsi parhaiten VR – ympäristössä ja mikä taas työasemaympäristössä?*
- *Minkä näet syynä tähän?*
- *Minkälaisia eroavaisuuksia näet näiden kahden ympäristön välillä?*
- *Kumpaa tapaa suosisit näiden tilanteiden selvittämiseen, VR vai työasema?*

Kysymysjoukkojen esittämisen vaihe käyttäjätutkimuksen aikana riippuu siitä, mikä demo-osio on ennen kysymyksiä suoritettu. Kolmen toisistaan eroavan kysymysjoukon esittämisen kriteerit ovat seuraavien sääntöjen mukaiset:

- Kysymysjoukko 1. esitetään ensimmäisenä aina kun on suoritettu maapallo tai pylväsdiagrammi ja graafi demo.
- Kysymysjoukko 2. esitetään kysymysjoukko 1. jälkeen aina kun demo on suoritettu kummassakin ympäristössä
- Kysymysjoukko 3. esitetään tutkimuksen lopuksi kaikkien demo-osioden suorittamisen jälkeen

### 6.1.2 Käyttäjätestin tehtävät

Käyttäjätestauksen sisällöksi on jokaiseen demojärjestelmään mietitty ennalta tehtäviä, joita käyttäjä pyydetään testin aikana tekemään tai esittämään vastaus. Tehtävien sisältö on jokaisessa osiossa säilytetty yksinkertaisena ja helposti ymmärrettävänä. Tehtävien tarkoituksena onkin lähinnä antaa motiivi järjestelmän sekä sen sisältämän datan hyödyntämiselle. Kaikille asetetuille tehtäville ei tarkoituksellisesti ole määritetty yksittäistä oikeaa vastausta, vaan käyttäjälle tarjotaan mahdollisuus omaan tulkintaan. Kysymysten vapaamuotoisuus myös toivottavasti saa käyttäjät pohtimaan vastauksia erilaisista näkökulmista ja etsimään niitä eri tavoin. Vapaamuotoisempien kysymysten lisäksi osaan kysymyksistä on etukäteen mietitty absoluuttinen oikea vastaus. Näiden kysymysten vastauksia ei kuitenkaan kirjata ja seurata käyttäjätutkimuksen aikana sen tarkemmin, vaan tutkimuksessa keskitytään käyttäjien käyttökokemuksiin perustuvien ajatuksien ja mielipiteiden kirjaamiseen.

Testaussuunnitelmassa esitellyn prosessin mukaisesti testin läpivieminen sisältää demo-osioden suorittamisen virtuaalitodellisuudessa sekä työasemaympäristössä. Kaikkiin kolmeen demo-osioon on luotuna kaksi eri tehtäväjoukkoa, jotka peruspohjaltaan ovat samat, mutta eroavat

kysymysten tarkemmassa muotoilussa. Ennen testin suorittamista jokaisessa demossa hyödynnettävänä oleva tehtäväjoukko on arvottu. Satunnaistamisen avulla tehtäväjoukkojen tulokset voidaan testien analysointivaiheessa arvioida riippumattomana suorituksen alaisesta ympäristöstä sekä testauksen aikaisesta oppimistekijästä. Käyttäjättestissä suoritettavien tehtäväjoukkojen tarkka sisältö esitellään seuraavaksi. Tarkemmat kysymyslomakkeet löytyvät tutkimuksen Liitteet-osiosta.

### **Maapallovisualisointi demon tehtävät**

#### Tehtäväjoukko 1.

- *Etsi maailmankartalta Australia.*
- *Etsi maailmankartalta Meksiko.*
- *Etsi maailmankartalta Suomi.*
- *Etsi kartalta jokin maa, josta on lähetetty yli 10 Twitter-viestiä.*
- *Etsi kartalta jokin maa, josta ei ole lähetetty yhtäkään Twitter-viestiä.*

#### Tehtäväjoukko 2.

- *Etsi maailmankartalta Kiina.*
- *Etsi maailmankartalta Saksa.*
- *Etsi maailmankartalta Yhdysvallat.*
- *Etsi Euroopasta jokin maa, josta on lähetetty yli 10 Twitter-viestiä.*
- *Etsi Afrikasta jokin maa, josta ei ole lähetetty yhtäkään Twitter-viestiä.*

### **Diagrammidemojen tehtävät**

#### Tehtäväjoukko 1. Pylväsdiagrammi

- *Mikä on auton 6. turvaluokitus?*
- *Miten auton 5. ominaisuudet mielestäsi suhteutuvat muiden diagrammissa olevien autojen ominaisuuksiin?*
- *Miten kuvailisit auton 3. ominaisuuksia?*

#### Tehtäväjoukko 1. Graafi

- *Mikä tieto graafissa on merkittävin tai suurin?*
- *Minkä tiedon painoarvo kasvaa?*
- *Onko graafin tiedon muutoksessa mielestäsi havaittavissa jotakin trendiä tai kaavaa?*

#### Tehtäväjoukko 2. Pylväsdiagrammi

- *Mikä henkilön henkilön 2 paino on?*
- *Miten kuvailisit henkilö 1 terveydentilaa annetun tiedon pohjalta?*

- *Miten henkilön 2 terveydentila mielestäsi suhteutuu muihin diagrammissa oleviin henkilöihin?*

#### Tehtäväjoukko 2. Graafi

- *Minkä tiedon painoarvo graafissa laskee?*
- *Mikä tieto graafissa on merkityksettömin tai pienin?*
- *Onko graafin tiedon muutoksessa mielestäsi havaittavissa jotakin trendiä tai kaavaa?*

Käyttäjätestin koko prosessi ja sen eri vaiheet on tarkemmin esitetty kaaviossa 1, jossa esitettyjen kysymys- ja tehtäväjoukkojen järjestys on yhä tarkemmin kuvattu tutkimuksen mahdollista toistamista varten.

## 6.2 Tulokset

Käyttäjätutkimus suoritettiin suunnitelman mukaisesti kymmenelle koehenkilölle. Testiin osallistujista kuusi oli miehiä ja neljä naisia. Osallistujia voidaan yleisesti kuvata nuoriksi aikuisiksi iän vaihteluvälin ollessa 18 – 39 vuotta. Testin osallistujista viisi kertoi omaavansa aikaisempaa kokemusta virtuaalitodellisuuden käyttöön liittyen. Aikaisempien virtuaalisten kokemusten selvityksen pohjalta yhtäkään osallistujaa ei kuitenkaan voida kategorisoida erityisen kokeneeksi virtuaalitodellisuuden käytön osalta. Vähäisten virtuaalitodellisuuteen pohjautuvien kokemusten vaikuttamattomuus huomattiin myös käyttäjätutkimuksen tietoja analysoitaessa. Aikaisemmalla kokemuksella tai osallistujan demograafisilla tiedoilla, ikä ja sukupuoli, ei ollut vaikutusta eri osioiden tulosten pisteyttämisessä. Tästä syystä osallistujiin liittyviä tekijöitä ei enää huomioida tulosten läpikäynnin tulevilla osioilla, vaan arvioista käytetään laskettuja kokonaiskeskiarvoja ja käyttäjien esittämät kommentit esitetään irrallisena osallistujan tiedoista. Tarkemmat käyttäjätutkimukseen annetut vastaukset ja arviot on esitetty osallistujakohtaisesti tutkimuksen Liitteet-osiossa.

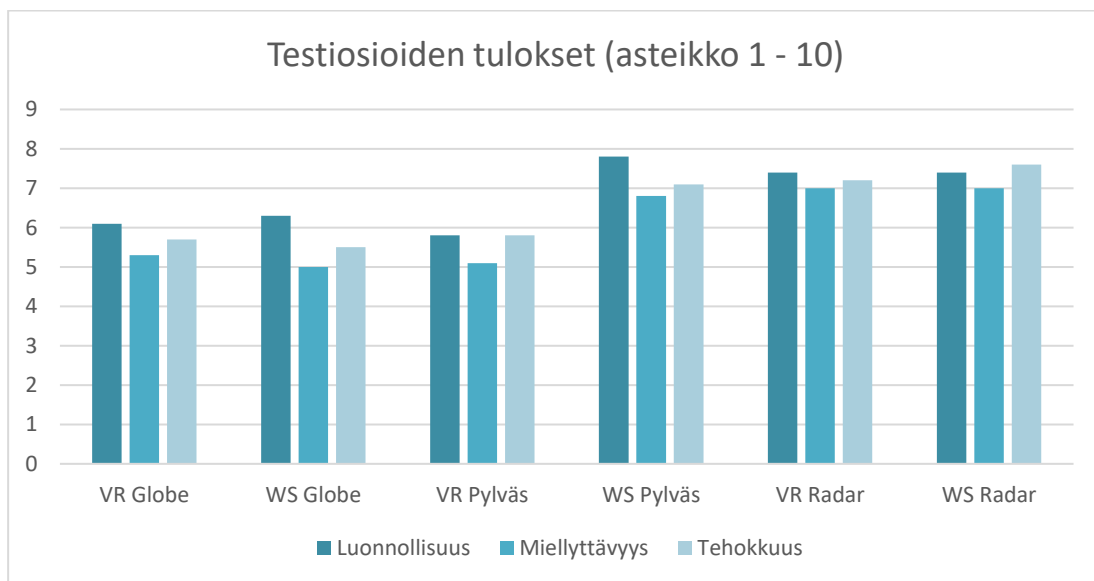
Seuraavaksi tutkimustulosten analysoinnissa käydään demo-osiokohtaisesti osallistujien arviot sekä vastaukset lävitse. Tämän jälkeen kappaleessa 6.2.2 käydään vielä lävitse käyttäjätutkimusten lopuksi esitettyjen koostavien kysymysten vastaukset ja esitetään pohdintoja vastausten taustoista.

### 6.2.1 Osioiden käyttökokemuksen arviot

Testi	Luonnollisuus	Miellyttävyyys	Tehokkuus
VR Maapallo	6,1	5,3	5,7
WS Maapallo	6,3	5	5,5
VR Pylväs	5,8	5,1	5,8
WS Pylväs	7,8	6,8	7,1
VR Graafi	7,4	7	7,2
WS Graafi	7,4	7	7,6

Taulukko 7. Käyttäjätestauksen kvantitatiiviset arvot taulukoituna.

Eri demo-osioiden käyttökokemusta arvioitiin jokaisen demon suorittamisen jälkeen yhdellä kuvailevalla ”Miltä osion käyttäminen tuntui”-kysymyksellä sekä kahdella kvantitatiivisella arviolla osion luonnollisuuden ja miellyttävyyden suhteen. Virtuaalitodellisuus ja työasema demon jälkeen osallistujaa pyydettiin lisäksi arvioimaan osioiden tehokkuutta kokemuksen pohjalta. Käyttäjätestauksessa saatujen kvantitatiivisten kysymysten tulosten keskiarvot on esitetty taulukossa 7 sekä pylväsdiagrammeina kaaviossa 2. Tuloksissa virtuaalitodellisuutta on kuvattu lyhenteellä VR sekä työasemaympäristöä lyhenteellä WS.



Kaavio 2. Käyttäjätestauksen tulokset kysymyskohtaisesti.

Näiden tietojen pohjalta voidaan sanoa, että Maapallo- sekä graafi-demojen osioiden välillä oli erittäin vähän hajontaa. Graafi-demon osioiden keskiarvojen ollessa lähes kaikkien arvojen osalta yhtenevät. Suurimmat eroavaisuudet esiintyivät pylväsdiagrammidemossa, jonka vastaukset kääntyivät selvästi suosimaan perinteistä työasemaa. Seuraavaksi käydään lyhyesti lävitse käyttäjien

antamat sanalliset kuvaukset jokaiseen osioon liittyen, jotka tarjoavat perusteita osioille annettujen arvioiden osalta.

### **1. Maapallovisualisointi virtuaalitodellisuudessa (VR maapallo)**

Lähes kaikki osallistujat ilmaisivat osion ja järjestelmän käytön olleen enemmän tai vähemmän kankeaa PS4-ohjaimen käytön vuoksi tai ohjaus ei vastannut sitä, mitä he olisivat odottaneet. Tämä demo oli myös ainoa, jossa muutamat osallistujat kokivat maltillista huonovointisuutta, jonka syyksi mainitsivat maapallon nopean kääntymisen. Moni osallistujista kuitenkin koki pään liikettä seuraavan kursorin käytön hyödylliseksi ja käyttöä nopeuttavaksi. Kaikki osallistujat eivät kuitenkaan pään kääntämistä halunneet tai huomanneet hyödyntää, vaan näkymän vaihtaminen toteutui pelkästään maapalloa ohjaimella kääntämällä. Katseenseurannan hyödyntämisen mahdollisuus voidaankin tulosten perusteella nähdä tässä positiivisena osatekijänä, joka vaikutti virtuaalitodellisuusympäristön hieman korkeampiin arvosanoihin tehokkuutta mitattaessa.

Suurin osa käyttäjistä koki osion käytön kokonaisvaltaisena kokemuksena, jossa he pääsivät lähemmäksi näkymää ja maapallon pintaa, mikä heidän mukaansa erityisesti helpotti maiden löytämistä työasemanäyttöön verrattuna. Tätä kokonaisvaltaisuuden tuntemusta voidaankin pitää yhtenä selittävänä tekijänä miellyttävyyden hieman paremmille arvioille työasemaympäristöön verrattuna.

### **2. Maapallovisualisointi työasemalla (WS-maapallo)**

Suurin osa osallistujista piti työasemalla toteutetun visualisoinnin kontrolleja intuitiivisena ja ne vastasivat osallistujien aikaisempia käyttötottumuksia. Aikaisemmat ja perinteisemmät käyttömallit selittävät työasemademon hieman korkeampia arviointeja käytön luonnollisuuden suhteen. Lähes puolet osallistujista pitivät työaseman käyttömallia kuitenkin hitaampana kuin virtuaalitodellisuuden toteutusta. Tämä johtui siitä, että käyttäjillä ei samaan tapaan kuin virtuaalitodellisuudessa ollut mahdollista valita maita hyödyntämällä omaa katsettaan. Lisäksi maapallon ”vedättäminen” hiirellä tuntui monesta hitaalta verrattuna ohjaimen kontrolleihin.

### **3. Pylväsdiagrammi virtuaalitodellisuudessa (VR-pylväs)**

Käyttäjätestin osallistujista vain kaksi henkilöä näki pylväsdiagrammitoteutuksen ainakin joiltain osin parempana virtuaalitodellisuudessa kuin työasemaympäristössä. Virtuaalitodellisuuden selkeästi huonompaa menestystä tässä demossa selittivät osallistujien vastausten pohjalta visualisoinnin heikompi hallittavuus sekä huonon resoluution takia sisällön



hahmottamisen heikentyminen. Osallistujat lisäksi kokivat, että pylväiden koon ja arvojen vertailu oli virtuaalitodellisuudessa hankalampaa. Tämä osittain johtunee siitä, että virtuaalitodellisuudessa osallistujat pääsivät jo liian lähelle itse visualisointia, jolloin perspektiivi sisältöön nähden heikkeni. Esimerkkinä tästä yksi osallistujista mainitsi kokeneensa olonsa ”hiireksi suurien pylväiden alapuolella”. Muutamat osallistujat myös kommentoivat, että he pitivät virtuaalitodellisuuden hyödyntämistä tällaisessa tilanteessa täysin turhana ja vain lisäepämukavuutta tuovana tekijänä päässä pidettävien lasien takia.

#### **4. Pylväsdiagrammi työasemalla (WS-pylväs)**

Pylväsdiagrammidemon käyttö työasemalla tuntui hieman odotetustikin osallistujien mielestä tavalliselta ja yllätyksettömältä. Muutamat osallistujat kokivat demon kontrollien hiirellä ja näppäimistöllä olevan hieman kankeita, mutta yleisesti demoa kuvattiin yksinkertaiseksi ja sen sisältöä helppolukuiseksi. Demon sisällön normaalius verrattaessa osallistujien aikaisempiin kokemuksiin datan hallinnasta varmasti vaikutti siihen, että demon käytöstä kirjatut kokemukselliset tiedot olivat hyvin yhtenäisiä ja osio saavutti kyselyssä selkeästi korkeammat arvot virtuaalitodellisuuteen verrattuna.

#### **5. Graafi virtuaalitodellisuudessa (VR-Graafi)**

Viimeisessä graafidemossa käyttäjien ei tarkoituksellisesti tarvinnut tehdä muuta kuin seurata datan muuttumista katseellaan. Tästä syystä kaikki osallistujat pitivätkin kokemusta helppona. Osa osallistujista koki näkymän seuraamisen virtuaalitodellisuudessa helpompana paremman kokonaiskuvan takia. Muutokset pistivät silmään paremmin ja päätä sekä katsetta ei tarvinnut erikseen käännellä tiedon ollessa virtuaalitodellisuudessa lähempänä käyttäjää. Yksi osallistujista kuitenkin mainitsi kokeneensa erityisesti virtuaalitodellisuudessa tunnelinäköä, jolloin hän koki tiedon tulleen liian lähelle häiriten keskittymistä muuta kuin yhteen tietopisteeseen kerrallaan. Samaan tapaan kuin maapallovisualisoinnin kohdalla, osallistujat eivät osanneet mainita suuria eroavaisuuksia, mutta kokivat tässäkin virtuaalitodellisuuden immersion parempana ja kokonaisuuden mielenkiintoisempana vaihtoehtona.

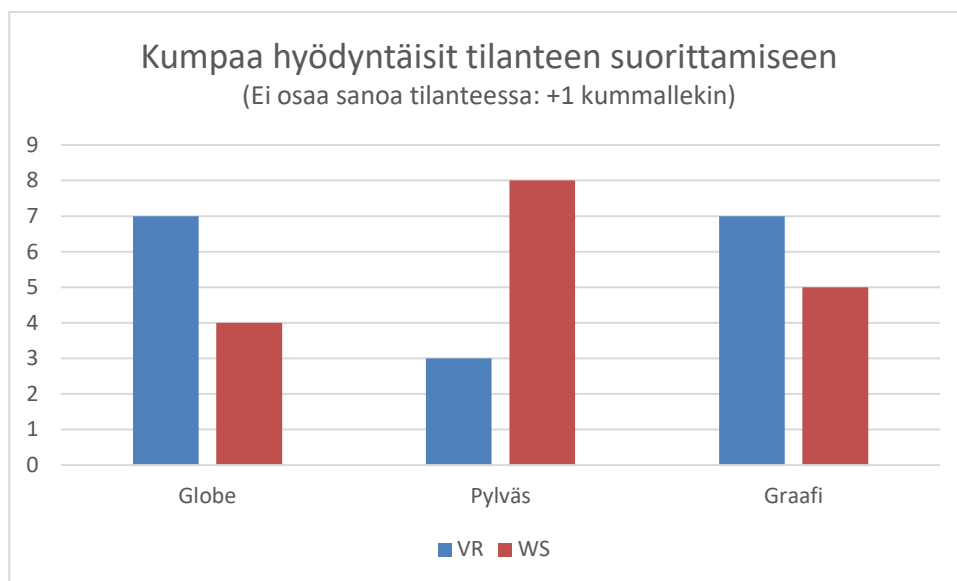
#### **6. Graafi työasemalla (WS-Graafi)**

Graafin osalta työasemaversiota pidettiin hyvin samanlaisena kuin virtuaalitodellisuuden toteutusta käyttäjälle tarjotun vuorovaikutuksen puuttuessa. Osallistujien vastauksissa työasemaversio keräsi positiivisia mainintoja siitä, että päässä ei tarvinnut pitää virtuaalilaseja

ja näkymä oli yksinkertaisuudessaan helppolukuinen. Kaksi käyttäjää mainitsi työasemaversiossa tarvittavan enemmän päänkääntelyä kaikkien muutosten seuraamiseksi, mikä työasemaversiossa näyttöä hyödynnettäessä onkin odotettua. Tehokkuutta lukuun ottamatta osiot saivatkin täysin saman keskiarvon luonnollisuuden ja miellyttävyyden osalta.

### 6.2.2 Loppukysymysten vastaukset

Käyttäjätutkimuksen loppukysymyksissä pyrittiin vielä summaamaan osallistujien ajatukset sen jälkeen, kun he olivat päässeet koittamaan kaikkia kolmea eri visualisointikonaisuutta. Kysymyksillä pyrittiin kartoittamaan ympäristöille ominaisia tekijöitä sekä miten osallistujat kokivat eroavaisuudet ympäristöjen välillä. Osallistujia pyydettiin myös kertomaan kumpaa ympäristöä demojen käyttämisen pohjalta suosisi erilaisissa käyttötapauksissa. Nämä osallistujien suositukset käyttöympäristöt on esitetty kaaviossa 3. Seuraavaksi osallistujien antamat vastaukset käydään tarkemmin lävitse kysymyskohtaisesti ja pohditaan vastauksien taustalla vaikuttavia tekijöitä.



*Kaavio 3. Osallistujien käyttäjätestissä suositukset käyttöympäristöt demokohtaisesti.*

#### 1. Mitkä demot toimivat parhaiten missäkin ympäristössä ja miksi?

Maapallo- ja graafi-demojen ympäristöjenvälisistä hyvin tasaisista ja virtuaalitodellisuuden jopa heikommista arvioista huolimatta osallistujat kallistuivat lopussa kuitenkin tukemaan virtuaalitodellisuuden vaihtoehtoa. Osa osallistujista kuvasi virtuaalitodellisuuden kokemuksen olleen useasti mielenkiintoisempi, intensiivisempi ja ympäristön kasvattaneen heidän keskittymiskykyään osion suorittamisen aikana sisällön tullessa niin sanotusti

”iholle”. Kuten osioiden tarkemmassa läpikäynnissäkin mainittiin, niin tämä kokonaisvaltaisemman kokemuksen tunne voidaan jälleen nähdä tekijänä, joka asetti käyttäjät arvioista huolimatta suosimaan virtuaalitodellisuuden toteutusta. Mahdollisesti myös uutuuden viehätys sai osallistujat painottamaan vastauksissaan mielenkiintoisempaa vaihtoehtoa.

Virtuaalitodellisuuden heikosta menestyksestä mainitsi kaksi osallistujaa, jotka olivat tutkimuksessa sitä mieltä, että virtuaalitodellisuus ei välttämättä soveltuisi minkään kolmen esityksen visualisointiympäristöksi, ja että se tuo mukanaan enemmän ongelmia kuin hyötyjä päässä pidettävien lasien ja heikomman näyttöresoluution muodossa. Tutkimuksen virtuaalitodellisuudessa toteutettujen demo-osioiden kielteisesti näkevät osallistujat olivat erityisesti sitä mieltä, että visualisointi tulisi olla tarkasti suunniteltu virtuaalitodellisuutta varten, jotta käytön hyödyt ylittäisivät ympäristölle ominaiset haitat.

Työasemaympäristön osioita kiiteltiin erityisesti niiden helposta lähestyttävyydestä ja työasemaympäristön tottumuksien ja normien mukaisuudesta. Mielenkiintoisena huomiona yhden osallistujan mukaan virtuaalitodellisuuden toteutukset eivät taas hänen mielestään vastanneet ympäristön mukaisia normeja, minkä hän koki käyttöä haittaavana tekijänä. Osallistuja ei kuitenkaan pystynyt näitä normeja nimeämään, mutta tämä oli häneltä osuva toteamus – mitä ovat virtuaalitodellisuuden interaktion normit ja onko niitä?

## **2. Minkälaisia eroavaisuuksia näit ja koit näiden kahden ympäristön välillä?**

Viisi osallistujaa mainitsi ympäristöjen eroavaisuuksissa virtuaalitodellisuuden olleen ajoittain tai koko ajan epämukava lasien painamisen tai yleisen päässä pitämisen takia. Kuusi osallistujaa myös mainitsi virtuaalitodellisuudessa häiritseviksi tekijöiksi erityisesti epätarkemman kuvanlaadun sekä hitaan kuvanpäivityksen. Nämä tekijät voidaankin erityisesti nähdä nykyisten virtuaalisten toteutusten haittatekijöinä, jos käytössä oleva laitteisto ja sen teho ei ole täysin ajanmukaista. Ennakkoon odotetusti järjestelmien toteutusvaiheessa työhön kohdistuneet laitteistokompromissit verottivat virtuaalitodellisuuden käyttökokemusta ja tämä näkyi osallistujien sanallisissa arvioissa. Kolme osallistujista kuitenkin mainitsi virtuaalitodellisuuden olleen demoissa parempi tapa päästä sisältöön kiinni ja poistaneen ulkopuoliset haittatekijät tehtäviä tehdessä. Varsinaisia merkittäviä eroavaisuuksia osallistujat eivät työasemaympäristön osalta maininneet muuta kuin käytön olleen normaalin näytön kautta vain hyvin perinteisen mukaista.

### **3. Kumpaa tapaa hyödyntäisit vastaavien tilanteiden suorittamiseen?**

Viimeisenä osallistujilta tiedusteltiin, kumpaa ympäristöä he suosisivat vastaavien tilanteiden suorittamisessa käyttäjätestin kokemuksien pohjalta. Vastausten hajautuminen oli kuvattu kaaviossa 3 ja sen pohjalta voitiin todeta osallistujien suosivan virtuaalitodellisuutta maapallo- ja graafi-demoa vastaavissa toteutuksissa ja pylväsdiagrammissa työasemaa. Usean osallistujan mielestä valinnan painotukseen vaikuttaisi käytössä oleva laitteisto. Laitteiston tulisi olla tässä tapauksessa tarpeeksi tehokas virtuaalimaailman pyörittämiseen saumattomasti ja lasien tulisi tuntua mukavilta päässä. Palautetta annettiin myös erityisesti siitä, että jotta virtuaalitodellisuuden käyttö tuntuisi merkitykselliseltä, tulisi toteutus ja moniulotteinen data pystyä tehokkaasti sitomaan osaksi virtuaalitodellisuuden toteutusta. Perinteisten visualisointimenetelmien suora siirtäminen virtuaalitodellisuuteen, kuten pylväsdiagrammin kohdalla, ei siis ole järkevää, vaan virtuaalitodellisuudessa visualisoinnin tulisi olla merkityksellinen, jotta käyttäjät ovat nykyisellään valmiita maksamaan virtuaalitodellisuudesta koituvan haitan. Useat osallistajat pitivät vastaamista kysymykseen myös hankalana, koska kokivat, että toteutus (esimerkiksi käyttäjän ja järjestelmän välinen vuorovaikutus) voi olla niin monimuotoinen ja järjestelmästä riippuvainen, mikä suoraan vaikuttaisi halukkuuteen hyödyntää jompaakumpaa ympäristöä.

## 7. YHTEENVETO

Tässä tutkimuksessa pyrittiin kartoittamaan vastauksia kahteen tutkimuskysymykseen:

- *Onko virtuaalitodellisuus soveltuva ympäristö Big Datan visualisoimiseen ja tehostaako se Big Dataksi luokiteltavien data joukkojen ymmärtämistä?*
- *Miten käyttäjien kokemus tiedon visualisoinnista eroaa virtuaalitodellisuuden ja työasemaympäristön välillä?*

Näiden kysymysten osalta tutkimuksessa tehtiin taustakartoitus Big Datan nykykäsitteestä sekä aikaisemmista virtuaalitodellisuuteen pohjautuneista Big Datan visualisointijärjestelmistä, joiden ominaisuuksia peilattiin nykyistä Big Datan käsitettä vasten. Taustatyön jälkeen tutkimuksessa pyrittiin luomaan visualisointeja, jotka toteuttaisivat Big Data käsitteen mukaisen määrittelyn ja niiden käyttökokemusta voitaisiin käyttäjätutkimuksessa verrata normaalia työasematoteutusta vasten. Testijärjestelmien toteutuksen jälkeen käytiin läpi Big Datan tuottamat haasteet erityisesti virtuaalitodellisuuteen pohjautuvassa järjestelmäkehityksessä ja esitettiin tästä johtuvat toteutukselliset puutteet. Testijärjestelmien kehityksen jälkeen niitä varten luotiin testisuunnitelma, jonka mukaisesti suoritettiin 10 osallistujan käyttäjätestaus tiedon visualisoinnin kokemusten selvittämiseksi yhtäläisen virtuaalitodellisuustoteutuksen ja työasemaversion välillä. Seuraavaksi käydään läpi yhteenveto tutkimuskysymysten osalta, jonka jälkeen esitellään tämän tutkimuksen pohjalta syntyneet jatkotutkimukselliset ajatukset.

Tutkielmassa tehdyn taustatyön pohjalta Big Datan yhdeksi suurimmista haasteista on nimetty Big Datan muuttaminen ymmärrettävään muotoon visualisointien avulla. Big Datan tehokkaan visualisoinnin ratkaisuja on pyritty hakemaan virtuaalitodellisuudesta toistaiseksi hieman heikoin lopputuloksin. Myös tässä tutkielmassa tehty selvitys virtuaalitodellisuuden hyödyntämisestä Big Datan visualisoinnissa toi esiin useita ongelmia huolimatta siitä, että tutkielmassa käytössä olevat resurssit ja laitteisto olivat rajalliset. Big Data asettaa jo itsessään merkittävän haasteen visualisoinnille sekä visualisoinnin järjestelmäkokonaisuudelle. Lisättäessä yhtälöön virtuaalitodellisuuden ympäristön, joka ei toistaiseksi vielä ole täysin lyönyt teknologiamarkkinoilla lävitse, haasteet moninkertaistuvat. Big Datan käsittely sekä virtuaalitodellisuuden hyödyntäminen vaativat kummatkin normaalista poikkeavan lähestymistavan, jolloin näiden yhteen liittäminen vaatii useiden haasteiden ratkaisemista aina laitteiston suorituskyvystä järjestelmän arkkitehtuuriin ja itse visualisoinnin käyttöliittymään sekä järjestelmän ja käyttäjän väliseen vuorovaikutustapaan.

Toinen suuri Big Datan hyödyntämiseen kohdistuva ongelma on sen rakenteen tuntemattomuus sekä ominaisuus muuttua ajan myötä. Tällöin dataa varten rakennetun visualisoinnin tulisi ainakin jossain määrin olla geneerinen toteutus, jossa visualisointi ei välittäisi miltään osin siihen syötetyn datan rakenteesta, mutta onnistuisi silti luomaan havainnollistavan visuaalisen kokonaisuuden. Virtuaalitodellisuutta hyödynnettäessä tällöin onkin vaara siitä, että visualisointi jää liian yleiselle tasolle, kuten tässä tutkielmassa sekä taustatutkimuksen mukaisissa aikaisemmissa virtuaalitodellisuuden visualisointijärjestelmissä, jolloin virtuaalitodellisuuden mahdollisuuksia ei päästä tehokkaasti käyttämään ja ympäristön hyödyntäminen menettää merkityksensä. Virtuaalitodellisuuden hyödyt voidaankin löytää lähinnä soveltuvan ja spesifin visualisoinnin käytöstä virtuaalitodellisuuden sisällä, jolloin visualisointi on luotu datavetoisesti eikä toisin päin. Tämä kuitenkin tarkoittaisi sitä, että visualisointi tulisi luoda erikseen jokaiselle datajoukolle, jolloin datajoukon sisältö tulisi tuntea, jotta tiedon pohjalta voitaisiin luoda tehokas visualisointi virtuaalitodellisuuteen. Tällöin kuitenkin visualisoinnille asetetut tavoitteet ja virtuaalitodellisuudelle suositellut ratkaisutavat sulkevat toisensa pois.

Tutkielmassa suoritettujen käyttäjätestien pohjalta voidaan todeta, että virtuaalitodellisuus tarjoaa hyviä mahdollisuuksia tehokkaiden visualisointiratkaisujen toteuttamiselle. Käyttäjätestauksessa huomioitiin, että useat käyttäjät kokivat virtuaalitodellisuuden visualisoinnit kokemuksellisempina sekä kokonaisvaltaisempina. Näiden tuntemusten pohjalta käyttäjät tunsivat, että he olivat paremmin osa visualisointia sekä pystyivät keskittymään sisältöön paremmin kuin työasemaympäristössä. Käyttäjätetit paljastivat myös sen, että perinteisiä, työasemaympäristöön luotuja visualisointeja, kuten diagrammit, ei koeta virtuaalitodellisuudessa merkityksellisinä. Työasemaympäristössä tutuksi tulleet visualisoinnit on luotu liiaksi noudattamaan kyseisen ympäristön normeja ja käyttökonventioita, mitkä eivät enää virtuaalitodellisuudessa toimi. Käyttäjätutkimuksen tulosten pohjalta osallistujat kokivat, että virtuaalitodellisuudessa toteutetun visualisoinnin tulisi hyödyntää ympäristön mahdollisuuksia esimerkiksi sitomalla data osaksi moniulotteisuutta tai hyödyntämällä tehokkaampaa järjestelmän ja käyttäjän välistä vuorovaikutusta. Tällöin visualisoinnin hyödyntämisellä on jokin merkitys ja ympäristö tehostaa tiedon sisäistämistä. Käyttäjätutkimuksen löydökset tukevatkin Big Datan visualisoinnin ongelman teoriaa, jossa viiden V:n määritteiden mukaisen Big Datan visualisoiminen loisi visualisoinnista liian geneerisen virtuaalitodellisuudessa, jolloin järjestelmän kehitykselle asetetut vaatimukset sekä hyödyt kumoutuvat. Tutkielman lähtökohteenä olleen geneerisesti Big dataa visualisoivan alustan luominen virtuaalitodellisuuteen saattaa vaarantaa käyttäjäkokemuksen ja käytön merkityksellisyyden, jolloin visualisointikokonaisuuden tulee toimiakseen olla erityisen hyvin luotu. Huomioitaessa virtuaalitodellisuuden laitteistolle sekä toteutukselle asettamat haasteet sekä käyttökokemuksen

huomattavasti tarkemman suunnittelemisen jäävät virtuaalitodellisuuden tuomat hyödyt tutkielman näkemyksien mukaan varsin pieniksi suhteutettuna siihen työmäärään, mitä tällaiset geneeriset visualisointialustat toimiakseen vaatisivat.

Jatkotutkimukselliset mahdollisuudet tutkielmassa käsiteltyjen aiheiden osalta ovat hyvin laajat. Tutkielman pohjalta pystyttiin vastaamaan vain siihen, että virtuaalitodellisuuden hyödyntäminen visualisoinneissa vaatii ympäristön tarpeet huomioon ottavan toteutuksen. Tutkittavaksi asiaksi kuitenkin jää mitä nämä tarpeet ja tekijät virtuaalisen ympäristön osalta voisivat olla. Tähän jatkotutkimuskysymykseen pohjautuen, yksi käyttäjätutkimuksen osallistujista mainitsi virtuaalitodellisuuden normit, jotka eivät hänen mukaansa tämän tutkielman kontekstissa täyttyneet. Ratkaisujen löytäminen virtuaalitodellisuuteen pohjautuvien visualisointien luomisessa voikin löytyä ympäristöön sovitettujen normien hyödyntämisestä, joita ei uuden teknologian osalta ole ainakaan vielä täysin muodostunut. Virtuaalitodellisuudessa hyödynnettävien laitteistojen tila on kuitenkin vielä nykyään kaikkea muuta kuin normalisoitunut markkinoiden hakiessa muotoaan. Tuotevalmistajat kehittävät omia tuotteitaan irrallaan toisistaan ja kilpailuvalttina hyödynnetään syötelaiteiden monimuotoisuutta sekä innovatiivisuutta. Sisällön tuottaminen virtuaalitodellisuuteen kuitenkin helpottuisi, jos virtuaalisesti luodun käyttöympäristön esitykselliset sekä vuorovaikutukselliset normit olisi käytön mukaisesti määritelty. Tällä hetkellä lähdettäessä luomaan uusia järjestelmiä virtuaalitodellisuuteen kehittäjä joutuu useiden eri valintatilanteiden eteen mietittäessä minkälaista syöte- ja liikkumistapaa tai ohjainta hän päättää hyödyntää. Yhtenä laajempänä käyttäjätutkimuksen mahdollisuutena olisikin tutkia useita eri vuorovaikutusmahdollisuuksia käyttäjän ja virtuaalitodellisuuden välillä ja kerätä tietoa eri konteksteihin sidottujen vuorovaikutustapojen toimivuudesta.



## LÄHDELUETTELO

- [Akerkar, 2013] Rajendra Akerkar, Big Data Computing. Taylor & Francis Group; 2013.
- [Beck, 2003] Michael Beck, Real-Time Visualization of big 3D City Models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol XXXIV-5/W10, Switzerland, February 2003.
- [Burdea & Coiffet, 2003] Grigore C. Burdea, Philippe Coiffet, Virtual Reality Technology. Wiley, New Jersey 2003.
- [Brown et al., 1995] Judith R. Brown, Rae Earnshaw, Mikail Jern, John Vince, Visualization: using computer graphics to explore data and present information. John Wiley & Sons, Inc., New York 1995.
- [Buyya et al., 2009] Rajkumar Buyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg..., Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems*. 25. 599-616. 10.1016/j.future.2008.12.001.
- [Card et al., 1999] Stuart K. Card, Jock Mackinlay, Ben Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers, 1999.
- [Chambers et al., 1983] John Chambers, William Cleveland, Beat Kleiner, Paul Tukey, Graphical Methods for Data Analysis. Wadsworth, 1983.
- [Chen & Liu, 2004] Keke Chen and Ling Liu. 2004. VISTA: validating and refining clusters via visualization. *Information Visualization* 3, 4, 257-270, December 2004.
- [Chen & Zhang, 2014] C.L. Philip Chen, Chun-Yang Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, *Information Sciences*, 275 (10), pp. 314-347, August 2014.
- [Chen, 2014] Keke Chen. 2014. Optimizing Star-Coordinate visualization models for effective interactive cluster exploration on Big Data. *Intell. Data Anal.* 18, 2, 117-136, March 2014.
- [Cox & Ellsworth, 1997] Michael Cox, David Ellsworth, Managing Big Data for scientific visualization. 1997.  
[https://www.researchgate.net/publication/238704525\\_Managing\\_big\\_data\\_for\\_scientific\\_visualization](https://www.researchgate.net/publication/238704525_Managing_big_data_for_scientific_visualization)
- [Cox et al., 1997] Donna J. Cox, Robert M. Patterson, Marcus L. Thiebaut, Virtual reality 3D interface system for data creation, viewing and editing. Patentti:  
<https://www.google.ch/patents/US6154723>

- [Cruz-Neira et al., 1993] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Surround Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, 135-142, 1993.
- [Donalek et al., 2014] Ciro Donalek, S.G. Djorgovski, Scott Davidoff, Alex Cioc ja muut. *Proceedings of 2014 IEEE International Conference on Big Data*, p 609-615, October 2014.
- [Du et al., 2016] Fei Du, A-Xing Zhu, Feng Qi, Interactive visual cluster detection in large geospatial datasets based on dynamic density volume visualization. *Geocarto International*, 31:6, 597-611, 2016.
- [Dutcher, 2014] Jennifer Dutcher, What is Big Data, 2014 Berkley School of Information. <https://datascience.berkeley.edu/what-is-big-data>.
- [Faber, 2013] Brenton Faber, Analytics, Big Data, and Technical Communication. In: *Association of Teachers of Technical Writing Conference*. Las Vegas, NV. March 2013.
- [Facebook, 2018] Facebook and Cambridge Analytica face class action lawsuit. *The Guardian*, 10.04.2018 <https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit>
- [Fernandez et al., 2014] Alberto Fernandez, Sara del Rio, Victoria Lopez, Abdullah Bawakid, Maria J. del Jesus... Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks. *WIRES Data Mining Knowl Discov*, 4: 380–409, 2014.
- [Few & Edge, 2007] Stephen Few, Perceptual Edge, Data visualization: past, present, and future. IBM Cognos Innovation Center, 2007.
- [Fisher, 2016] Danyel Fisher, Big data exploration requires collaboration between visualization and data infrastructures. *HILDA '16 Proceedings of the Workshop on Human-In-the-Loop Data Analytics*. San Francisco, California, July 2016.
- [Foo et al., 2009] Jung-Leng Foo, Thom Lobe, Eliot Winer, A Virtual Reality Environment for Patient Data Visualization and Endoscopic Surgical Planning. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, vol 19, supplement 1, 211-218, 2009.
- [Graph And Chart, 2018] ProSource Labs, Graph And Chart. <https://www.assetstore.unity3d.com/en/#!/content/78488> Katsottu 15.04.2018
- [Haber & McNabb, 1990] Robert B. Haber, David A. McNabb, Visualization Idioms: A Conceptual Model for Scientific Visualization Systems. In: G.M. Nielson, B. Shriver, L. Rosenblum

- (eds.) *Visualization in Scientific Computing*, *IEEE Computer*, 74{93. IEEE Computer Society Press Los Alamitos, CA, USA (1990)
- [Hadi et al., 2014] Hiba Jasim Hadi, Ammar Hameed Shnain, Sarah Hadishaheed, Azizahbt Haji Ahmad, Big Data and Five V's Characteristics. Big Data and Five V's Characteristics. *Proceedings of IRF International Conference*, 1st November 2014, Tirupati, India.
- [Harrison & Römhild, 2007] Chris Harrison, Christoph Römhild, Bible Cross-References. <http://www.chrisharrison.net/index.php/Visualizations/BibleViz>
- [IBM, 2013] IBM: The Four V's of Big Data, 2013. <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- [InfoVis, 2018] InfoVis Wiki, [http://www.infovis-wiki.net/index.php?title=Visualization\\_Pipeline](http://www.infovis-wiki.net/index.php?title=Visualization_Pipeline) Katsottu 15.04.2018
- [Inselberg & Dimsdale, 1990] Alfred Inselberg, Bernard Dimsdale. 1990. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Vis'90: Proceedings of the IEEE Visualization '90*, pages 361–370. IEEE Computer Society, 1990
- [Intel, 2013] Intel, Big Data Visualization: Turning Big Data Into Big Insights. <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/big-data-visualization-turning-big-data-into-big-insights.pdf>
- [International Data Corporation, 2014] International Data Corporation, Data Growth Business Opportunities, and the IT Imperatives. April 2014, <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.html>
- [Johnson et al., 2005] Chris Johnson, Robert Moorhead, Tamara Munzner, Hanspeter Pfister, Penny Rheingans, Terry S. Yoo. 2005. NIH-NSF Visualization Research Challenges Report. Los Alamitos, California: IEEE.
- [Kandogan, 2000] Eser Kandogan, Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions. *Proceedings of the IEEE information visualization symposium, late breaking hot topics*, p. 9–12, 2000.
- [Kandogan, 2001] Eser Kandogan, Visualizing multi-dimensional clusters, trends, and outliers using star coordinates, *Proceedings of ACM SIGKDD Conference*, 107–116, 2001.
- [Keahey, 2013] T. Alan Keahey, Using visualization to understand Big Data. Technical Report, IBM Corporation, 2013, 1-16.
- [Kehrer et al., 2012] J. Kehrer, R.N. Boubel, P. Filzmoser, H. Piringer, A Generic model for the integration of interactive visualization and statistical computing using R. *Conference on Visual Analytics Science and Technology (VAST)*, IEEE 2012 p. 233-234

- [Khan & Khan, 2011] Muzammil Khan, Sarwar Shah Khan, Data and Information Visualization Methods and Interactive Mechanisms: A Survey, *International Journal of Computer Applications*, 34(1), 1-14, 2014.
- [Kim, 2005] Gerard Jounghyun Kim, Designing Virtual Reality Systems, The Structured Approach. Springer London, London 2005.
- [Klein & Gorton, 2015] John Klein, Ian Gorton, Runtime Performance Challenges in Big Data Systems. *Proceedings of the 2015 Workshop on Challenges in Performance Methods for Software Development*. January 2015.
- [Latin Square, 2018] Latin Square Design. <http://www.statisticshowto.com/latin-square-design/> Katsottu 15.04.2018
- [Laney, 2001] Doug Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety. META Group, February 2001.
- [Leap Motion, 2018] Leap Motion. <https://www.leapmotion.com/> Katsottu 15.04.2018.
- [Likert, 1932] Rensis Likert, A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55. 1932
- [Long & Linsen, 2011] Tran Van Long, Lars Linsen, Visualizing high density clusters in multidimensional data using optimized star coordinates. *Computational Statistics* 26:655, December 2011.
- [Marcos et al., 2013] Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A. S. Netto..., Big Data Computing and Clouds: Trends and Future Directions. *Journal of Parallel and Distributed Computing*, Volumes 79-80, 3-15, May 2015.
- [Miksch & Aigner, 2014] Silvia Miksch, Wolfgang Aigner, A matter of time: applying a data-users tasks design triangle to visual analytics of time-oriented data. *Comp Graph*, 2014, 38:2, 86-90.
- [Miranskyy et al., 2016] Andriy Miranskyy, Abdelwahab Hamou-Lhadj, Enzo Cialini, Alf Larsson, Operational-Log Analysis for Big Data Systems: Challenges and Solutions. *IEE Software*, vol. 3, issue 2, pp. 52-59. 2016.
- [Monino, 2016] Jean-Louis Monino, Data Value, Big Data Analytics and Decision-Making. *Journal of the Knowledge Economy*, 1-12, 2016.
- [Moran et al., 2015] Andrew Moran, Viljay Gadepally, Matthew Hubbell, Jeremy Kepner, Improving Big Data visual analytics with interactive virtual reality. *IEEE High Performance Extreme Computer Conference (HPEC)*, pp. 1–6, 2015.

- [Moreland, 2013] Kenneth Moreland, A Survey of Visualization Pipelines, *IEEE Transactions on visualizations and computer graphics*, vol. 19, no. 3, March 2013.
- [Muller & Schumann, 2003] Wolfgang Muller, H. Schumann, Visualization method for time dependent data—an overview. *Proceedings of the 2003 Winter Simulation Conference*, vol. 1. IEEE; 2003.
- [Neves & Bernardino, 2015] Pedro Neves, Jorge Bernardino, Big Data Issues. *IDEAS'15 Proceedings of the 19<sup>th</sup> International Database Engineering & Applications Symposium*, p 200-201, Yokohama, Japan, July 2015.
- [Ngrain, 2013] Ngrain, 3 reasons why “visualization” is the biggest “V” for Big Data. <https://actuariatfinance.wordpress.com/2014/12/01/3-reasons-visualization-biggest-v-big-data/>
- [OSTP, 2012] Office of Science and Technology Policy 2012, Obama administration unveils "Big Data" initiative: Announces \$200 million in new R&D investments. [http://www.whitehouse.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)
- [Olshannikova et al., 2015] Ekaterina Olshannikova, Aleksandr Ometov, Yevgeni Koucheryavy, Thomas Olsson, Visualizing Big Data with Augmented and Virtual Reality: Challenges and Research Agenda. *Journal of Big Data*, Vol. 2, 22, 01.10.2015.
- [Onemilliontweetmap, 2018] #Onemilliontweetmap. <https://onemilliontweetmap.com/> Katsottu 15.04.2018.
- [Playfair, 1786] William Playfair, The Commercial and Political Atlas. London, 1786.
- [Plugfelder & Helmut] Plugfelder, Ehren Helmut, Big Data, Big Questions. *Communication Design Quarterly* 1, no. 4, 18-22, 2013.
- [Rajaraman, 2016] V. Reson Rajaraman, Big Data Analytics. *Resonance*, August 2016, Volume 21, Issue 8, 695 – 716.
- [Rijmenami, 2016] Mark van Rijmenam, Why The 3V's Are Not Sufficient To Describe Big Data. <https://datafloq.com/read/3vs-sufficient-describe-big-data/166>. 2016
- [Rubio-Sanchez & Sanchez, 2014] Manuel Rubio-Sanchez, and Alberto Sanchez, Axis Calibration for Improving Data Attribute Estimation in Star Coordinates Plots. *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 2013-2022, December 31 2014.
- [Shneiderman, 1992] Ben Shneiderman, Tree visualization with tree-maps: 2-d space-filling approach. *Journal ACM Transactions on Graphics*, 11,1, 92-99, January 1992.

- [Shvachko et al., 2010] Konstantin Shvachko, Hairong Kuang, Sanjay Radia, Robert Chansler, The Hadoop Distributed File System. *MSST '10 Proceedings of the 2010 IEEE 26<sup>th</sup> Symposium on Mass Storage Systems and Technologies (MSST)*.
- [Sucharitha et al., 2014] Venkataramana Sucharitha, S.R. Subash, P. Prakash, Visualization of Big Data: Its Tools and Challenges, *International Journal of Applied Engineering Research*, 9(18), pp. 5277-5290, 2014.
- [Sutherland, 1968] Ivan E. Sutherland, A Head-mounted three dimensional display. *Proceedings of the AFIPS Fall Joint Computer Conference*, Thompson Books, 757-764, Washington D.C, 1968.
- [Szewczyk et al., 2004] Rovert Szewczyk, Joseph Polastre, Alan Mainwaring, David Culler, Lessons from a sensor network expedition. *European Workshop on Wireless Sensor Networks*, 307-322. 2004.
- [Teng et al., 2015] Pengju Teng, Hongjun Li, Xiaopeng Zhang, Survey on Visualization Layout for Big Data. International Conference on Intelligent Science and Big Data Engineering, Intelligence Science and Big Data Engineering. *Big Data and Machine Learning Techniques* pp 384-394, October 2015.
- [Teräs & Raghunathan, 2015] Marko Teräs, Shriram Raghunathan, Big Data Visualisation in Immersive Virtual Reality Environments: Embodied Phenomenological Perspectives to Interaction. *International Journal of Soft Computing*. 05. 1009-1015, August 2015.
- [Tolle et al., 2005] Gilman Tolle, Joseph Polastre, Robert Szewczyk ja muut, A macroscope in the redwoods. *Sensys '05 Proceedings of the 3rd international conference on Embedded Networked Sensor Systems*, 51-63. November 2005, San Diego, California, USA.
- [Turner et al., 2014] Vernon Turner, J. F. Gantz, D. Reinsel, S. Minton, The digital universe of opportunities: rich data and the increasing value of the internet of things. IDC EMC, 2014. <https://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm>
- [Twitter Stream API, 2018] Twitter Streaming API. <https://developer.twitter.com/en/docs/Katsottu> 15.04.2018.
- [Unity, 2018] Unity 3D, cross-platform game engine developed by Unity Technologies. <https://unity3d.com/> Katsottu 15.04.2018.
- [Vicon, 2018] Vicon, Motion Capture Systems. <https://www.vicon.com/> Katsottu 15.04.2018

- [Vo et al., 2011] Huy T. Vo, Jonathan Bronson, Brian Summa, Joao L.D Comba..., Parallel visualization on large clusters using MapReduce. Large Data Analysis and Visualization (LDAV),
- [Wang et al., 2015] Lidong Wang, Guanghui Wang, Cheryl Ann Alexander, Big Data and Visualization: Methods, Challenges and Technology Progress. *Digital Technologies*. Vol. 1, No. 1, 2 33-38, 2015.
- [Ware, 2004] Colin Ware, Information Visualization – Perception for Design. Morgan Kaufmann, April 2004.
- [Watson, 2014] Hugh J. Watson, Tutorial: Big Data Analytics: Concepts, Technologies, and Applications. *Communications of the Association for Information Systems*: Vol. 34 , Article 65, 2014.
- [World Political Map, 2018] Kronnect, World Political Map – Globe Edition. Editor Extension. <https://www.assetstore.unity3d.com/en/#!/content/41890> Katsottu  
15.04.2018
- [Wu, 2017] Hao Wu, Big Data Management the Mass Weather Logs. Smart Computing and Communication. SmartCom 2016. *Lecture Notes in Computer Science*, vol 10135. 2017
- [Yahoo, 2016] Yahoo Confirms at Least 500 Million Accounts Were Hacked. [Fortune.com/2016/09/22/yahoo-hack](https://fortune.com/2016/09/22/yahoo-hack)



## LIITTEET

### Taustatietolomake

Käyttäjätestin aineistoa käytetään vain tämän tutkimuksen käyttötarkoituksia varten. Käyttäjätesti on mahdollista keskeyttää missä vaiheessa tahansa.

Ikä: \_\_\_\_\_

### Sukupuoli:

☐ Mies

☐ Nainen

### Oletko käyttänyt virtuaalitodellisuus-laseja aikaisemmin:

☐ Kyllä

☐ En

Jos **kyllä**, voitko lyhyesti kertoa aikaisemmista kokemuksistasi:

---

---

---

---

---

## Käyttäjätutkimuksen tehtävät:

### Maapallovisualisointi

#### Tehtäväjoukko 1.

Tehtävä 1.	Etsi maailmankartalta Australia.
Tehtävä 2.	Etsi maailmankartalta Meksiko.
Tehtävä 3.	Etsi maailmankartalta Suomi.
Tehtävä 5.	Etsi kartalta Euroopasta jokin maa, josta on lähetetty yli 10 Twitter-viestiä.
Tehtävä 6.	Etsi kartalta Aasiasta jokin maa, josta ei ole lähetetty yhtäkään Twitter-viestiä.

#### Tehtäväjoukko 2.

Tehtävä 1.	Etsi maailmankartalta Kiina.
Tehtävä 2.	Etsi maailmankartalta Saksa.
Tehtävä 3.	Etsi maailmankartalta Yhdysvallat.
Tehtävä 5.	Etsi Aasiasta jokin maa, josta on lähetetty yli 10 Twitter-viestiä.
Tehtävä 6.	Etsi Etelä-Amerikasta jokin maa, josta ei ole lähetetty yhtäkään Twitter-viestiä.

#### Tehtäväjoukko 1. Pylväsdiagrammi

Tehtävä 1.	Mikä on auton 7. turvaluokitus?
Tehtävä 2.	Miten auton 5. ominaisuudet mielestäsi suhteutuvat muiden diagrammissa olevien autojen ominaisuuksiin?
Tehtävä 3.	Miten kuvailist auton 3. ominaisuuksia?

#### Tehtäväjoukko 1. Graafi

Tehtävä 1.	Mikä tieto graafissa on merkittävin tai suurin?
Tehtävä 2.	Minkä tiedon painoarvo kasvaa?
Tehtävä 3.	Onko graafin tiedon muutoksessa mielestäsi havaittavissa jotakin trendiä tai kaavaa?

#### Tehtäväjoukko 2. Pylväsdiagrammi

Tehtävä 1.	Mikä henkilön henkilön 2 paino on?
Tehtävä 2.	Miten kuvailisit henkilö 1 terveydentilaa annetun tiedon pohjalta?
Tehtävä 3.	Miten henkilön 2 terveydentila mielestäsi suhteutuu muihin diagrammissa oleviin henkilöihin?

## Tehtäväjoukko 2. Graafi

Tehtävä 1.	Minkä tiedon painoarvo graafissa laskee?
Tehtävä 2.	Mikä tieto graafissa on merkityksettömin tai pienin?
Tehtävä 3.	Onko graafin tiedon muutoksessa mielestäsi havaittavissa jotakin trendiä tai kaavaa?

## Käyttäjä tutkimuksen haastattelukysymyslomakkeet

### Kysymysjoukko 1. Yksittäisen demon jälkeen

Miltä osion käyttäminen tuntui?									
Kuinka luonnolliselta järjestelmän käyttö tuntui?									
1	2	3	4	5	6	7	8	9	10
Epäluonnolliselta									Luonnolliselta
Kuinka miellyttävältä järjestelmän käyttö tuntui?									
1	2	3	4	5	6	7	8	9	10
Epämiellyttävältä									Miellyttävältä

### Kysymysjoukko 2. Diagrammi/Graafi/Aikajana demon jälkeen

Kuinka tehokkaalta järjestelmän käyttö virtuaalitodellisuudessa tuntui?									
1	2	3	4	5	6	7	8	9	10
Epätehokkaalta									Tehokkaalta
Kuinka tehokkaalta järjestelmän käyttö tietokoneen näytöllä tuntui?									
1	2	3	4	5	6	7	8	9	10
Epätehokkaalta									Tehokkaalta

### Kysymysjoukko 3. VR + Työasema demon jälkeen

Mikä näistä kolmesta demosta: Pylväsdiagrammi, graafi tai Maapallo toimi mielestäsi parhaiten tässä VR ympäristössä ja mitkä työpöytäympäristössä?

Minkälaisia eroavaisuuksia näet näiden kahden ympäristön välillä?

Kumpaa tapaa suosisit näiden tilanteiden selvittämiseen VR vai työasema?

## Käyttäjätutkimuksen tulokset ja taulukot

### Osallistujat

Osallistuja	1	2	3	4	5
ikä	27	24	39	25	39
Sukupuoli	Nainen	Nainen	Mies	Mies	Mies
Aikaisempi kokemus	Ei	Kyllä	Kyllä	Kyllä	Kyllä
Kokemus kuvaus		Kokeilin myyntidemossa. Lasien käyttö sujui silloin ongelmitta. Kokeillut myös lasia museossa jossa katsoin videota.	Kaverilla pelannut pariin kertaan ja nähnyt pari demoa. Lisäksi messudemo koittanut	Paria demosovellusta testannut. Pieniä pelejä, leikkiä ja kokeilua	Pelasin messuilla HTC-Vivella, ja nyt uskon VR:n lyövän läpi
6	7	8	9	10	
25	N/A	26	26	18	
Nainen	Mies	Mies	Mies	Nainen	
Kyllä	Ei	Ei	Ei	Ei	
Lyhyesti kokeillut muutamia demoja. Kokeillut ajaa autoa viven ja ratin kanssa.					

### Vastaukset käyttäjäkohtaisesti

#### Luonnollisuus

Testi	1	2	3	4	5	6	7	8	9	10
VR Globe	7	7	8	4	5	7	7	5	4	7
WS Globe	9	9	7	5	8	6	7	3	3	6
VR Pylväs	4	10	4	6	5	4	8	3	8	6
WS Pylväs	8	10	6	8	7	9	9	6	7	8
VR Radar	3	10	7	9	8	9	4	7	10	7
WS Radar	4	10	7	9	8	7	5	7	10	7

#### Miellyttävyyys

Testi	1	2	3	4	5	6	7	8	9	10
VR Globe	4	7	7	2	5	7	6	6	4	5
WS Globe	5	9	4	5	7	3	6	3	3	5
VR Pylväs	4	10	3	2	4	5	7	3	8	5
WS Pylväs	7	9	3	5	6	9	9	6	7	7
VR Radar	2	9	6	8	8	9	4	8	9	7
WS Radar	4	10	5	9	8	6	5	6	10	7

**Tehokkuus**

<b>Testi</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
VR Globe	5	6	7	3	6	7	6	6	5	6
WS Globe	3	7	6	5	8	8	7	3	3	5
VR Pylväs	3	9	6	4	6	7	6	3	8	6
WS Pylväs	8	8	5	6	7	10	8	6	6	7
VR Radar	3	9	5	9	8	10	5	7	9	7
WS Radar	5	10	7	9	8	9	5	6	9	8